THE EVOLUTION OF MORPHOLOGICAL COMPLEXITY IN CYANOBACTERIA

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(*Dr. sc. nat.*)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Bettina E. Schirrmeister

aus

Deutschland

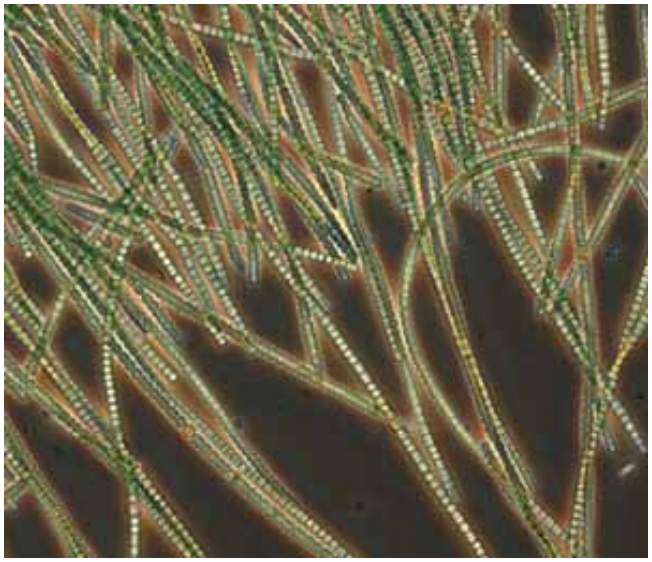
Promotionskomitee

Prof. Dr. Homayoun C Bagheri (Leiter der Dissertation)

Prof. Dr. Lukas Keller

Prof. Dr. Kentaro K Shimizu

Zürich, 2012



BETTINA E. SCHIRRMEISTER

DISSERTATION

The Evolution of Morphological Complexity in Cyanobacteria

Author:
Bettina E. SCHIRRMEISTER

Supervisor:
Prof.Dr. Homayoun BAGHERI

February 21, 2012

Contents

List of Figures

List of Tables

"Freudig war ſeit vielen Jahren
Eifrig ſo der Geiſt beſtrebt,
Zu erforſchen, zu erfahren,
Wie Natur im Schaffen lebt.
Und es iſt das ewig Eine,
Daſſ ſich vielfach offenbart;
Klein das Groſſe, groſſ das Kleine,
Alles nach der eigenen Art,
Immer wechſelnd, feſt ſich haltend,
Nah und fern, und fern und nah,
ſo geſtaltend, umgeſtaltend —
Zum Erſtaunen bin ich da."

Johann Wolfgang von Goethe

(1749-1832)

Summary

The evolution of morphological complexity in cyanobacteria

Cyanobacteria are one of the oldest and most diverse prokaryotic phyla known. They are assumed to beresponsible for what is known as the "Great Oxydation Event"(GOE) more than 2.4 billion years ago. Oldest fossil remains are known from 2.0 billion year old rocky cherts. By that time, they already exhibited the majority of their diverse morphological characteristics. This morphological variety that in many cases resembles simple algal forms, led to a misinterpretation in the 19th century of cyanobacteria as part of the plant kingdom. Although their lack of many eukaryotic features suggested a bacterial relation in the early 20th century, the debate on how to classify cyanobacteria was ongoing till the 1960s. With the development of DNA sequencing techniques, molecular phylogenetic classification accelerated. The first phylogenetic trees of cyanobacteria using 16S rRNA sequences were built in the 1990s. The expansion of molecular sequencing techniques resulted in new opportunities for cyanobacterial classification on the genome level. But these techniques are still progressing, few genomes of cyanobacteria are known and the full cyanbacterial diversity still remains to be explored. With this dissertation, I intend to use 16S rRNA sequences to unfold cyanobacterial diversity as it is known today. Using a carefully chosen subset I reconstruct the evolution of cyanobacterial morphotypes. Via phylogenetic methods, I hope to answer questions on the origin of multicellularity and get insights regarding prerequisites for the evolution of aerobic life.

In **CHAPTER I "Copy number variation and sequence conservation of ribosomal genes in cyanobacteria"**,

I explored known genome data of cyanobacteria. Multiple gene copies are known to have effects on phenotypes in eukaryotes. Recently, it has been shown that cyanobacterial evolution is not strongly directed towards genome reduction. Therefore, multiple gene copies might play a more important role than previously thought. Various highly conserved gene copies of ribosomal genes have been found to affect growth rates and fast adaptation to changing environments in bacteria. In this study, we identified 44 genes including ribosomal genes occurring as multiple highly conserved gene copies, in cyanobacteria. The distribution of these gene copies was compared to levels of cell differentiation. Increased ribosomal gene copy numbers were strongly correlated with species capable of terminal differentiation. Furthermore, 16S rRNA showed extreme sequence conservation within genomes. This statement held also for the overall 16S rRNA variation if compared to other eubacterial phyla. It seems that (i) increased rRNA copies might confer an advantage associated with terminal differentiation, (ii) multiple 16S rRNA copies are highly conserved as a consequence of concerted evolution in combination with purifying selection, and (iii) evolutionary rates are very low in cyanobacteria. Furthermore, phylogenetic results were compared to phylogenomic analyses. It seems that the usage of 16S rRNA sequences as a taxonomic marker in phylogenetic classification would lead to useful results.

In **CHAPTER II "The origin of multicellularity in cyanobacteria"**, I explored the evolution of multicellularity using phylogenetic tools. The phylogenetic position of cyanobacteria among other eubacterial phyla was investigated. Cyanobacteria have a monophyletic origin within the Eubacteria. For a unicellular cyanobacterial species with a simplified photosynthetic apparatus, named *Gloeobacter violaceus* outgroup testing confirmed a position close to the eubacterial outgroup. In order to identify the complete known diversity in cyanobacetria, I reconstructed a phylogenetic tree of 1,220 cyanobacterial strains using 16S rRNA sequences. From this large dataset a subset was chosen to represent the complete diversity of this phylum. Using this cyanobacterial subset phylogenetic trees were reconstructed using maximum likelihood and Bayesian inference. On 10,000 trees from the Bayesian analyses the evolution of multicellularity was reconstructed using methods of maximum likelihood and parsimony. The results indicate that, (i) cyanobacteria share a unicellular most recent common ancestor, (ii) multicellularity evolved very early during Earth history leading to the majority of extant cyanobacterial species, including four of five morpholohical sections, and (iii) multicellularity was not only lost several times throughout cyanobacteria history, but also regained at least once.

In **CHAPTER III "Evolution of cyanobacterial morphotypes: Taxa required for improved phylogenomic approaches"**, I discuss strategies for sampling cyanobacterial species for genome sequencing. This should be done in a way that presents the entire diversity for the cyanobacterial phylum. Based on the large phylogenetic tree reconstructed in chapter II, we point out previous sampling biases towards marine unicellular bacteria, and elaborate on species choice for future genome studies which would enable better analyses on the evolution of cyanobacteria. Furthermore we review the kinds of phylogenomic approaches that would be applicable.

In **CHAPTER IV "Evolution of multicellularity coincided with diversification of cyanobacteria and the**

Great Oxidation Event", I try to elucidate the time of the origin of multicellularity, and investigate potential associations with the "Great Oxidation Event"(GOE). Cyanobacterial fossils are found frequently in cherts from 2.0 billion years ago and younger. Furthermore, it is assumed that cyanobacteria caused the rise of atmospheric oxygen more than 2.4 billion years ago. Using Bayesian inference we tried to reconstruct the origin and early evolution of cyanobacteria that subsequently had a strong influence on Earth's history. The combination of diverse calibration priors led to results that indicate several distinct statements on early cyanobacterial history. (i) Cyanobacteria have clearly originated before the GOE, (ii) multicellularity coincides with the beginning of the GOE, in addition to increased diversification rates in the phylum, and (iii) three of the major cyanobacterial clades, comprising 80-85% of known strains, originated shortly after the alteration of Earth's atmosphere.

Clearly, this study would ideally be extended to a genome level. But the phylogenetic results presented here, reveal interesting aspects of early cyanobacterial evolution. Multicellularity might have had various advantages for cyanobacterial growth and distribution and could have affected oxygen production more than 2.4 billiion years ago. The incorporation of more precise geochemical and further fossil data in combination with a phylogenomic analysis could help to further resolve the timing and the nature of the the early diversification of cyanobacteria and its effects on the young Earth.

Zusammenfassung

Die Evolution morphologischer Komplexität in Cyanobakterien

Cyanobacterien gehören zu den ältesten und morphologisch vielfältigsten Prokaryoten. Sie sind verantwortlich für das sogenannte "Grosse Oxidations Ereignis" (GOE;*Great Oxidation Event*) vor etwa 2,4 Milliarden Jahren. Die ältesten Fossilfunde sind in 2,0 Milliarden Jahren alten Gesteinsschichten entdeckt worden und umfassen bereits die meisten der noch heute vorhandenen Erscheinungsformen. Aufgrund dieser morphologischen Vielfalt, welche oft simplen Algen sehr ähnlich ist, wurden Cyanobakterien im 19. Jahrhundert, dem Reich der Pflanzen zugeordnet. Obwohl bereits zu Beginn des 20. Jahrhunderts, dass Fehlen vieler eukaryotischer Merkmale, eine enge Verwandtschaft zu Bakterien nahelegte, waren sich Wissenschaftler bis Mitte 20. Jahrhundert über die phylogenetische Zuordnung von Cyanobakterien unschlüssig. Durch die Entwicklung neuer Methoden zur Sequenzierung von Nukleinsäuren, wurde die phylogenetische Klassifizierung enorm vorangetrieben. Erste phylogenetischen Bäume der Cyanobakterien basieren auf 16S rRNA Sequenzdaten und wurden in den 90er Jahren des letzten Jahrhunderts rekonstruiert. Die Erweiterung der Sequenzierungstechniken ermöglichte eine phylogenetische Klassifizierung auf molekularer Ebene der Cyanobakterien auf der Genom Ebene. Diese Techniken stehen jedoch noch am Anfang. Bisher sind wenig Genome von Cyanobakterien sequenziert und die tatsächliche Diversität dieses Stammes lässt sich nur erahnen. Mit einer sorgfältig ausgewählten Zusammensetzung verschiedener Cyanobakterienarten möchte ich die Evolution der cyanobakterieller Morphotypen rekonstruieren. Unter Verwendung phylogenetischer Methoden erforsche ich die Entstehung der Vielzelligkeit in Cyanobakterien und die Vorraussetzungen für die Evolution

aeroben Lebens auf der Erde.

In **KAPITEL I "Unterschiede in der Anzahl und starke Konservierung von ribosomalen Genen in Cyanobakterien"**, habe ich bekannte Genomdaten von Cyanobakterien untersucht. Von Eukaryoten ist bekannt, dass das Vorkommen mehrerer Genkopien starke Effekte auf den Phänotypen eines Organismus haben kann. Kürzlich wurde widerlegt, dass Evolution in Cyanobakterien in Richtung einer Genomreduktion verläuft. Daher könnte es sein, dass multiple Genkopien eine wichtigere Rolle spielen als bisher angenommen. Beispielsweise, ermöglicht eine höhere Anzahl an konservierte rRNA Genen, beschleunigtes Wachstum in Bakterien und eine schnellere Anpassung an veränderte Umweltbedingungen. In dieser Studie haben wir 44 Gene, inklusive ribosomaler Gene, identifiziert, die in mehrfacher Ausführung in Cyanobakterien vorkommen. Die Anzahl der Genkopien wurde mit dem Grad der jeweiligen Zelldifferenzierung verglichen. Vermehrte ribosomale Genkopien korrelierten mit Arten, die terminale Zelldifferenzierung aufweisen. Des Weiteren war die 16S rRNA Gensequenz erstaunlich stark konserviert. Diese Aussage liess sich auch für den ganzen Stamm der Cyanobakterien, im Vergleich mit anderen Bakterienstämmen, nachweisen. Es scheint, dass in Cyanobakterien, (i) erhöhte Ribosom Kopien einen Vorteil, im Zusammenhang mit terminaler Zelldifferenzierung, aufweisen, (ii) 16S rRNA Genkopien stark konserviert sind, aufgrund von gemeinsamer Evolution (*concerted evolution*) und reinigende Selektion (*purifying selection*), und (iii) Evolutionsraten sehr gering sind. Zudem wurden die phylogenetischen Ergebnisse mit phylogenomischen verglichen und es scheint, dass die Verwendung von 16S rRNA Gensequenzen, als taxonomisches Mittel zur phylogenetischen Analyse, zu durchaus brauchbaren Ergebnissen führt.

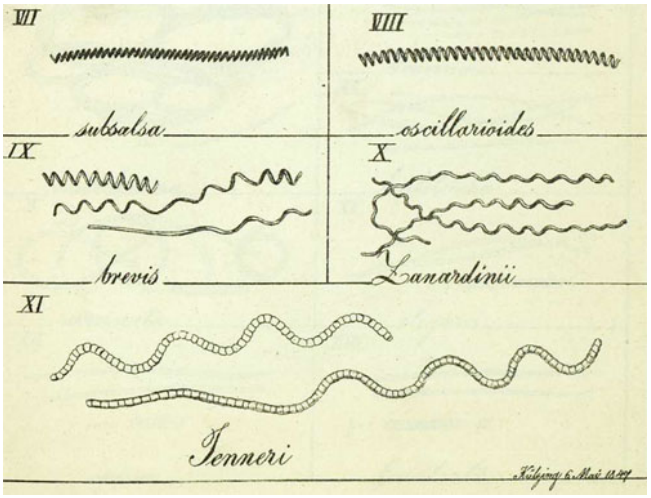
In **KAPITEL II "Der Ursprung der Vielzelligkeit in Cyanobakterien"**, habe ich, mit Hilfe von phylogenetischen Methoden die Entwicklung der Vielzelligkeit in Cyanobakterien untersucht. Die phylogenetische Position in Beziehung zu den Eubakterien wurde bestimmt. Cyanobakterien haben einen monophyletischen Ursprung innerhalb der Eubakterien. Das Testen der Aussengruppe bestätigte die nahe Position *Gloeobacter violaceus* und *Synochococci* arten, welche von thermalen Quellen isoliert wurden, zur Aussengruppe. Um die gesamte, vorhandene Vielfalt der Cyanobakterien einschätzen zu können, rekonstruierten wir einen phylogenetischen Baum basierend auf 16S rRNA Daten, der 1,220 verschiedene Cyanobakterien-taxa beinhaltet. Von diesem Baum wurde eine Teilmenge an Cyanobakterien sorgfältig ausgewählt, um die gesamte vorhandene Variation dieses Phylums zu berücksichtigen. Mit dieser Teilmenge an Cyanobakterien wurden phylogenetische Bäume rekonstruiert unter Verwendung von *Maximum Likelihood* und Bayesscher Analysen. An mehr als 10,000 phylogenetischen Bäumen der Bayesschen Analyse, wurde die Evolution von Vielzelligkeit rekonstruiert. Verwendet wurden *Maximum Likelihood* und Parsimonie Methoden. Die Ergebnisse deuten darauf hin, dass (i) alle Cyanobakterien von einem gemeinsamen einzelligen Vorfahren abstammen, (ii) sich sehr früh in der Geschichte der Cyanobakterien eine vielzellige Linie entwickelt hat, von der heute die Mehrzahl aller vorhandenen morphologischen Gruppen abstammen, und (iii), dass Vielzelligkeit während der Evolution der Cyanobakterien, mehrmals verloren gegangen ist und sich mindestens einmal wieder entwickelt hat.

In **KAPITEL III "Evolution verschiedener Morphotypen in Cyanobakterien: Benötigte Taxa für verbesserte phylogenomische Studien"**, erörtere ich Strategien für eine verbessertes Auswahl von Cyanobakterienarten für phylogenomische Studien. Ausgewählte Arten zur Genomsequenzierung sollten dazu beitragen, die Vielfalt dieses Phylums zu repräsentieren. Anhand des grossen phylogenetischen Baumes, welcher in Kapitel II rekonstruiert wurde, weisen wir darauf hin, wie die vorhandenen Cyanobakterien Genome stark in Richtung einzellige, marine Cyanobakterien verschoben ist. Des Weiteren erörtern wir, aus welchen Cyanobakteriengruppen man Arten für zukünftige Genom Projekte idealerweise wählen sollte, um die Evolution dieses Stammes adäquat ergründen zu können. Zusätzlich diskutieren wir wie phylogenomische Ansätze aussehen könnten.

In **KAPITEL IV "Dìe Evolution Vielzelligkeit überschnidt sich mit erhöhter Diversifizierung von Cyanobakterien und dem "Grossen Oxidations Ereignisses"**, versuche ich herauszufinden, wann Vielzelligkeit entstanden ist und ob es Hinweise auf einen möglichen Zusammenhang mit dem "Grossen Oxidations Ereignis"(GOE) gibt. Fossile Cyanobakterien wurden bisher in 2,0 Milliarden Jahre alten und jüngeren Gesteinsschichten gefunden. Darüber hinaus wird angenommen, dass Cyanobakterien für die Anreicherung von Sauerstoff vor über 2.4 Milliarden Jahren verantwortlich sind. Mithilfe Bayesscher Analyse haben wir den Ursprung und die frühe Evolution von Vielzelligkeit, zu rekonstruiert. Eine Entwicklung, die nachträglich einen starken Einfluss auf die Evolution der Erde, gehabt hat. Die Kombination verschiedener Kalibrierungspunkte hat zu Ergebnissen geführt, die auf verschiedene eindeutige Aussagen hinsichtlich der Evolution von Cyanobakterien, hinweisen. (i) Cyanobakterien sind eindeutig vor dem GOE entstanden, (ii) die Entwicklung von Vielzelligkeit überschneidet sich mit dem GOE und mit erhöhter Diversifikiationsrate in diesen Stamm, und (iii) drei Gruppen, die heute 70-80% der Cyanobakterien ausmachen, entwickeln sich kurz nachdem die Erdatmosphäre mit Sauerstoff angereichert wurde.

Idealerweise würde diese Studie zu einem Phylogenomischen Projekt erweitert. Dennoch ermöglichen die hier präsentierten phylogenetischen Ergebnisse interessante Aussagen, hinsichtlich der frühen Entwicklung von Cyanobakterien. Vielzelligkeit könnte verschiedene Vorteile gegenüber anderen Bakterien gehabt und zu einer weiteren Ausbreitung von Cyanobakterien geführt haben, mit entsprechenden Folgen für die globale Sauerstoff Produktion. Der Einbezug von geochemischen und fossilen Daten in Kombination mit phylogenomischen Analysen könnte weitere Einblicke in die frühe Diversifikation von Cyanobakterien und mögliche Folgen für die junge Erde, liefern.

INTRODUCTION



A brief history of cyanobacterial phylogenetics

Author:

Bettina E. SCHIRRMEISTER

For more than 2000 years, living organism were considered static and not related to each other. They were positioned on a ladder-like frame, a so called "scala naturae". One of the first to recognize the changing character of nature was the French naturalist Jean Baptiste Lamarck. His theory would be continued and expanded by naturalists of the 19th century, such as Alfred Wallace and Charles Darwin. The latter presented in 1859 a theory of evolution driven by natural selection. Consequently, the concept of a tree for the representation of species relatedness would be developed. Although phylogenetic trees were not used widely at the beginning, the concept gained popularity from the 1960s onwards. For cyanobacteria (blue-green algae) the usage of new technologies would lead to a reclassification. Traditionally, blue-green algae and bacteria had been grouped within the plant kingdom. Within this group the relationship of blue-green algae was debated. Reasons for these taxonomic conflicts were strong morphological similarities of some blue-green algae to some algal growth forms, and the occurrence of chlorophyll within cells. Chlorophyll is a color pigment found in plants, but not in other bacteria. On the other hand they lack features like sexual reproduction and a nuclei, negative characteristics that are shared exclusively with bacteria. When in 1938, it became clearer that bacteria should rather be classified in a separate kingdom, the debate where to place cyanobacteria would go on until 1970, when it would be established that cyanobacteria are part of the prokaryotes. The biological evolution of the morphological variety that caused debates for decades, still needs to be explored. Many of these cyanobacterial forms have been present more than 2.0 billion years ago. Furthermore, cyanobacteria are assumed to have raised oxygen levels more than 2.4 billion years ago, during the “Great Oxidation Event” (GOE). Therefore, understanding the evolution of cyanobacterial morphotypes would result in important insights on the early evolution of Earth and the beginnings of respiratory life.

On Phylogenetic trees

> *"the analogy of a branching tree, as the best mode of representing the natural arrangement of species and their successive creation"*
> **Alfred R. Wallace** (1855)

With this statement Alfred R. Wallace (1823-1913) is among the first to suggest a branching tree for the description of species relatedness. The idea for such a tree derives from the attempt to explain variation between different species.

History of a tree (1739-1900)

"Every species has come into existence coincident both in space and time with a pre-existing closely allied species." **Wallace** (1866)

Wallace assumed that some sort of correlation existed between different species. The idea that species were related to each other and change over time developed in the 19th century. Before, nature would have been described exclusively on the basis of similarities. The predominant opinion was that nature was static, and organisms could be placed on a ladder according to their level of complexity. A so called *scala naturae* was devised, where "wise man" (*Homo sapiens*) would be seated at the very top [1, 2] (Figure 0.1). This hierarchical framework for the description of living organisms derived from Aristotle's (384-322 B.C.) "*Historiae Animalium*" [3]. Aristotle placed animals according to functional and structural complexity on a ladder where the level of perfection rises towards the top. This was a classification system that was in no respect based on an evolutionary thought. Different groups of organisms were assumed to be independent of each other and constant in appearance. This picture of nature was basically maintained

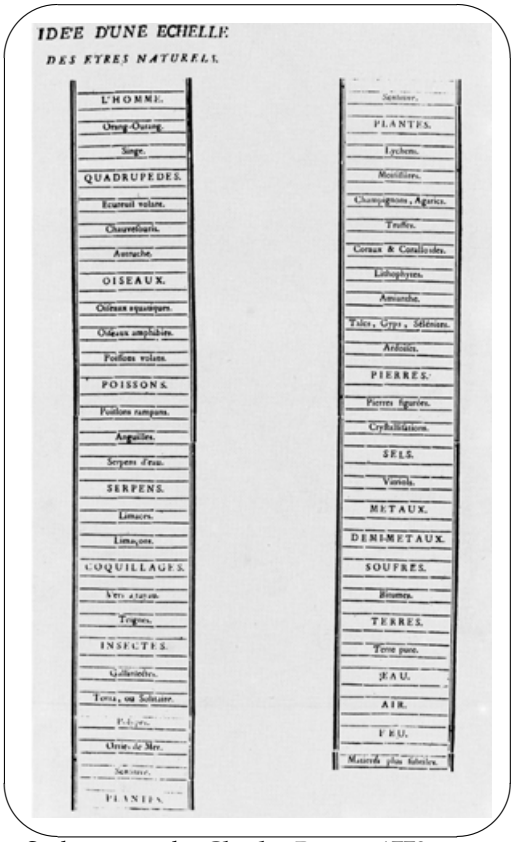


Fig. 0.1. Scala naturae by **Charles Bonnet** 1779

to the end of the 18th century. Linné (1707-1778) implemented this hierarchical framework in the book "*Systema Naturae*" (1739) which was one of the first comprehensive approaches to classify species [4]. Around 1780, the adequacy of the *scala naturae* to represent the order of the natural system was questioned, but suitable alternatives were missing [5]. The German poet and naturalist Johann Wolfgang von Goethe (1749-1832) and the French naturalist Jean Baptiste Lamarck (1744-1829) developed the idea that living organisms are not static and seem to change over time [6]. A first tree-like sketch can be found in Lamarck's "Philosophie Zoologique" [7], even though it is still based on a ladder-like structure of the living world. The idea that species were changeable would be carried on by the French naturalist E. Goeffroy Saint-Hilaire (1772-1844), but at the same time be denied by many contemporaries, one of them being his French colleague George Cuvier (1769-1832) [4, 8]. Alfred R.Wallace supported Lamarck's and Hilaire's idea of a transmutation of species. In 1855, he came to the conclusion that species are created on the type of pre-existing ones. An idea he deduced from observations of island species, such as found on Galapagos archipelago [9].

> *"A country having species, genera, and whole families peculiar to it, will be the necessary result of its having been isolated for a long period, sufficient for many series of species to have been created on the type of pre-existing ones, which, as well as many of the earlier formed species, have become extinct, and thus made the groups appear isolated."* **Wallace** (1855)

Impressed by Charles Lyell's "The principles of Geology" [10] Wallace stated that gradual geological changes might have altered the environments in ways that could have led to the disappearance of some species. And although he still believed that species evolved towards a higher degree of perfection he admitted that such a progression could not be observed within organismic groups. In this respect he mentioned mollusca, radiata and fishes where fossil representatives appeared more advanced than living species.

The year Lamarck published his provocative book on the heritability of acquired physiological changes is also the birth year of English naturalist Charles R. Darwin (1809-1882), who 50 years later would propose the concept of evolution driven by natural selection . Independently of Wallace, Darwin would draw similar conclusions by reviewing observations which he made during his voyage on the HMS Beagle (1831-1836). In his "Notebook B - the transmutation of species" (1837-1838) he drew what became one of the most famous sketch of an evolutionary tree, introduced by the words *"I think"* (Figure 0.2).

In the mid 19th century Alfred R. Wallace and Charles R. Darwin had come up with similar thoughts regarding the causality of observed variation in species. Since his return from his Beagle-voyage Darwin had worked on a concept to explain the diversity of species. In 1858, Lyell and J. D. Hooker therefore suggested a joint publication of the conclusions made by Darwin and Wallace under the title *"On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection"* to the Linnean society [11]. This subsequently encouraged Darwin to finally publish an abstract "On the Origin of Species by Means

of Natural Selection", which was intended to be part of his big book "Natural selection" [12]. In the introduction Darwin would acknowledge that others potentially have come to similar conclusions around that time, for "a naturalist who combines such facts as the affinity of species, geographic distributions, and geological data might conclude a non random descent of species" [12].

"Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) requires extinction. Thus between A & B immense gap of relation. C & B the finest gradation, B & D rather greater distinction. Thus genera would be formed. — bearing relation to ancient types. — with several extinct forms for if each species an ancient (1) is capable of making 13 recent forms, twelve of the contemporarys must have left no offspring at all, so as to keep number of species constant. — "

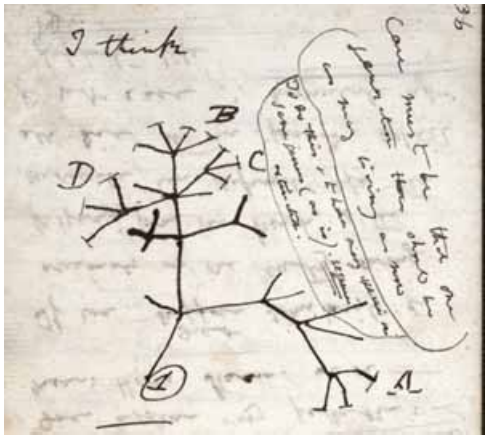


Fig. 0.2 Sketch of evolutionary tree by **Charles .R. Darwin** 1837

A distinct observation Darwin made in his book is the idea of natural selection as the motor of evolution. For the following 70 years, the mechanism of natural selection would face critics from various fields [13, 14, 15]. But, most of Darwin's ideas found strong support from scientists such as English anatomist Thomas H. Huxley (1825-1895), American botanist Asa Gray (1810-1888) and German naturalist Ernst H. Haeckel (1834-1919). Subsequently, the thought of a "tree-like" concept started to become more prominent [12, 6, 2] (Figure 2). Species were no longer assumed to be static but credited to evolve in time, possibly with natural selection as the driving force. Organisms were assumed to be related to each other and to descend from pre-existing species. Therefore, living species could potentially be pictured like the tips of branches on a tree which converged in a stem that represented some sort of common ancestor. Such a tree can be found in the book, *"Generelle Morphologie der Organismen"* (general morphology of organisms) from 1866, where Ernst Haeckel intended to combine the fields of Botany and Zoology on the basis of evolution as described by Darwin, and to introduce theories of the latter to the German public. Haeckel was delighted with Darwin's publication [6] and would call Darwin together with Goethe and Lamarck the fathers of evolution [16]. The idea of a branching tree for the description of the relationships of species appeared logical to Haeckel.

> *"der Entwurf der organischen Stammbäume, obwohl gegenwärtig noch äusserst schwierig und bedenklich, wird meines Erachtens die wichtige und interessante Aufgabe für die Morphologie der Zukunft bilden"* (The scheme of organismic phylogenies, although quite difficult and questionable at present, will as to my judgment be an important and interesting challenge for the Morphology of the future.) **Haeckel** (1866)

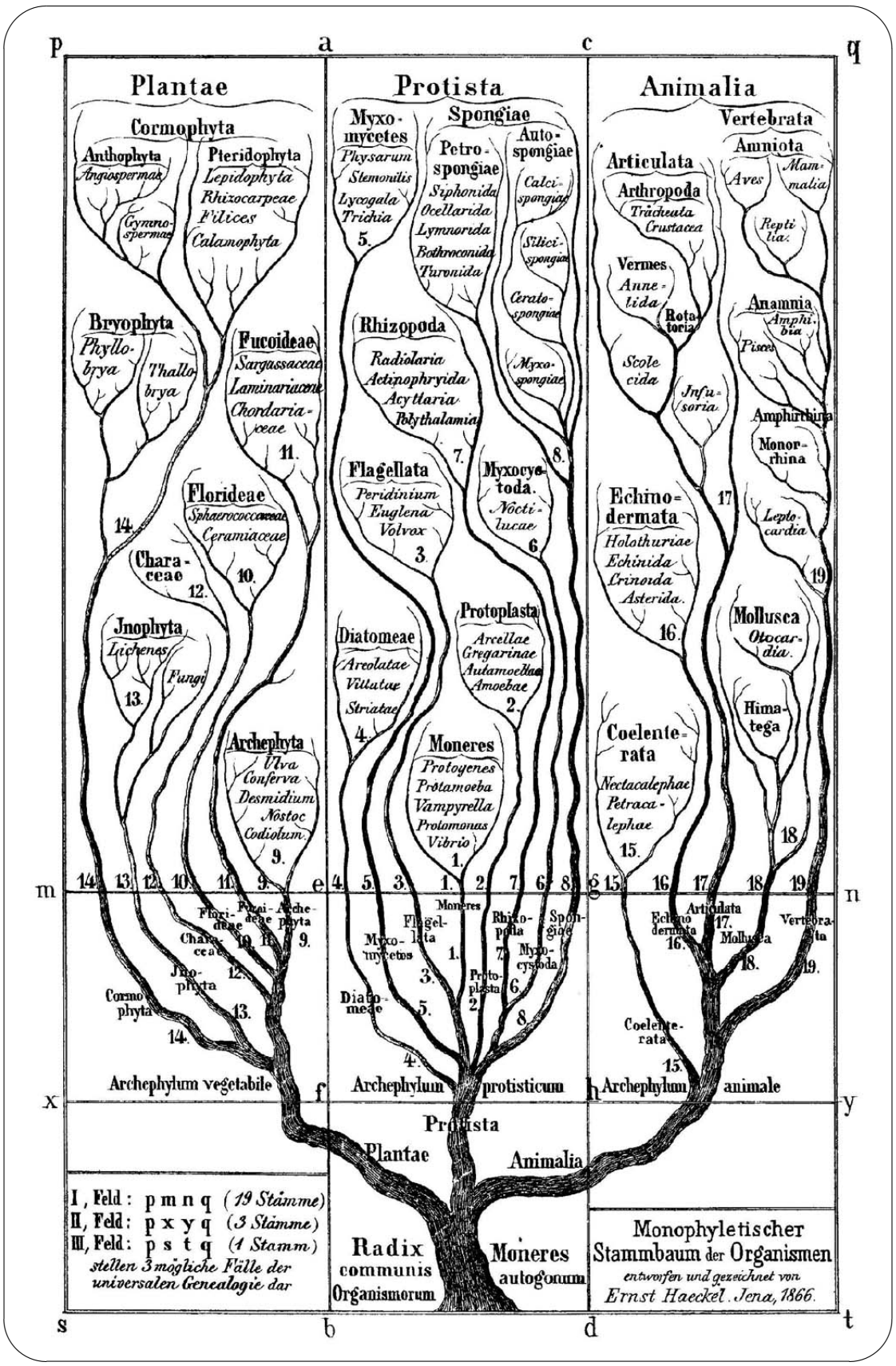


Fig. 0.3. "Monophyletischer Stammbaum der Organismen" by **Ernst H.P.A. Haeckel** 1866

Inspired by Darwin's theory Haeckel intended to find the true monophyletic genealogical tree of life [6]. Although he was technically a zoologist, he nevertheless tried to combine all organism (including plants) with regard to the "Deszendenz-Theorie" (that all organisms originate from a single common ancestor). Reason was, as he claimed, that most zoologists and especially botanists at that time ignored general and comprehensive questions, possibly due to a lack of overall knowledge. In Haeckel's opinion this results in books "rich in paper and devoid of mind" [6]. In his pursuit to find a universal explanation for the morphologies observed in nature, Haeckel created what is cited today as the first phylogenetic tree of life [16] (Figure 3). Throughout his life Haeckel would try to continue this pursuit, to find the universal tree which incorporated all forms of organisms. Darwin's idea of evolution gained wide acceptance in the following years. However, the thought of natural selection as a driving force remained debated in the scientific community [14, 17, 15]. The following years would be described as the "eclipse of Darwinism", a phrase originally formulated by Julian Huxley and subsequently used to describe the effects of anti-Darwinian thought around 1900 [18, 17, 15]. Furthermore, the idea of a phylogenetic tree that connects species according to their level of relatedness would lose relevance for some time.

Mendelians and biometricians (1900-1920)

In 1900, the laws of heritability published originally in 1859 by the Austrian monk Gregor J. Mendel (1822-1884) were rediscovered — independently — by Hugo de Vries and Carl Correns. The consequence would be a split of evolutionary biologists into Mendelians and biometricians. Biometricians, with popular members such as the English mathematician Karl Pearson and W. F. R. Weldon, would try to find an algorithm that could explain gradual, phenotypic Darwinian evolution led by natural selection ("gradualism"). Mendelians, including English geneticist William Bateson, argued for discrete genetic variation as it was supposedly observed in nature. A variation that changed erratically according to laws of inheritance, not in agreement with natural selection ("saltationism") [14]. The controversy between these schools would last for many years. By 1918, genetic experiments such as conducted by Thomas H. Morgan (1866-1945) would show how mutations would increase variation within a species, and how natural selection would select for the persistence of beneficial mutations ([19] p.188-190; [14] p.120-125). It appeared Mendel's laws on inheritance and Darwin's selection did not contradict each other, but were rather complementary ([14]p.130). In 1918, the English statistician R. A. Fisher (1890-1960) would publish the mathematical basis to explain the continuous variation as arguemented by Darwinists as a result of Mendelian inheritance [20].

Population genetics and the evolutionary synthesis (1918-1950)

The combination of "Darwinism, Mendelism and biometry" ([14]p.131) to understand evolution would be expressed by the new school of population genetics based on works of R. A. Fisher [21], S. Wright (1889-1988) [22] and J. B. S Haldane (1892-1964) [23]. Their theories focused on the interaction of alleles and genes in a

population, and showed that mendelian genetics is indeed consistent with the concept of natural selection and gradualism. The subsequent attempts to combine microevolutionary results from population genetics with macroevolutionary observations from palaeontology led to the emergence of the "Modern Evolutionary Synthesis" around 1936-1947 [13]. The intention of this area was to use evolutionary theory to present a synthesis between various biological schools of thought. The evolutionary synthesis was catalyzed by the ideas of A. R. Wallace and August Weismann (1834-1914) [24], who rejected the ideas of acquired inheritance ([25];[26] p.252). Prior to this, Lamarck's inheritance of acquired changes, so called ' soft inheritance" was still considered a possible mechanism of evolution [13, 18, 24]. Prominent scientist of the "Modern Synthesis of Evolution" were, among others, Julian Huxley, Ernst Mayr, G. G. Simpson, and Theodosius Dobzhansky (1900-1975). Dobzhansky, an Ukrainian geneticist and evolutionary biologist worked on wild populations of fruitflies. In 1937, he and his colleague Sturtevant published one of the first genetic trees based on gene arrangement differences on the 3rd Chromosome of *Drosophila pseudobscura* and sister species [27] (Figure 4).

"a comparison of the different gene arrangements in the same chromosome may, in certain cases, throw light on the historical relationships of these structures, and consequently on the history of the species as a whole". **Dobzhansky and Sturtevant** (1937)

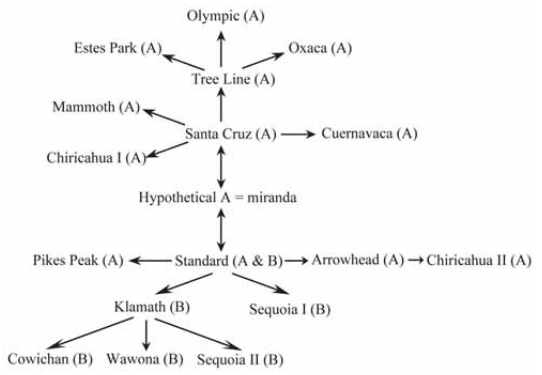


Fig. 0.4 Phylogeny of Drosophila pseudobscura strains as indicated by gene arrangements adapted from Dobzhansky and Sturtevant [27] (right)

Cladistics

In the 1930s biochemists such as Marcel Florkin (1900-1979) and Ernest Baldwin (1909-1969) investigated possibilities to use comparative biochemistry for the inference of phylogenetic relations of animals [28]. In 1970, Florkin explicates the intention of his 1944 book "L'èvolution biochimique" as a call for comparative biochemistry and its usage in phylogenetics. In 1950, Willi Hennig (1913-1976) a German entomologist published "Grundzüge einer Theorie der phylogenetischen Systematik" (main features of a theory regarding the phylogenetic classification). In his book he introduced a phylogenetic classification system based on species relationship [29]. This theory would replace classification systems which were focused solely on morphological similarities without the incorporation of evolutionary relation. Hennig's framework is based on monophyly, which describes groups of species sharing a distinct common ancestry, an idea that Ernst Haeckel had also proposed in 1866 [6, 29]. In 1966, a translated version "Phylogenetic systematics" appeared and is retrospectively viewed as a turning point for phylogenetic classification [30, 31]. The work of Hennig introduced cladistics, a method which groups species that share a common ancestor into "clades", a term introduced by Huxley in 1958 [31].

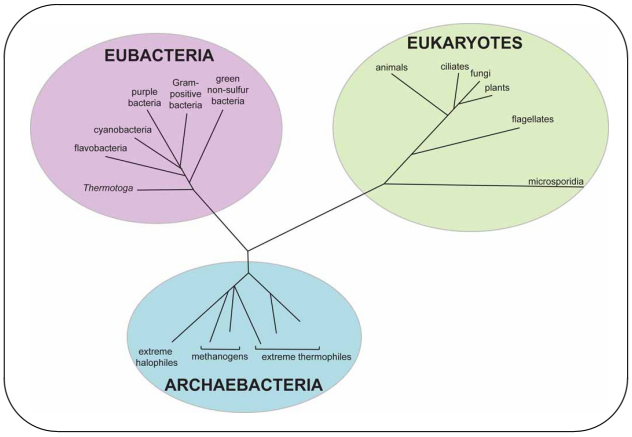


Fig. 0.5. Universal phylogenetic tree determined from 16S rRNA sequence comparison adapted from **Woese** *et al.* 1987

Opposed to the cladistic approach suggested by Hennig, taxonomists relied on phenetics, a classification system based on morphological similarities regardless of phylogenetic relation of species. In 1963, English microbiologist Peter H. A. Sneath and Austrian biostatistician Robert R. Sokal introduced the field of numerical taxonomy based on the concepts of phenetics [32]. The method relies on cluster analyses of observable distances between species. At the same time, critics on the weaknesses of phenetic methods accumulated [33, 34, 35]. With increasing computational power, cladistical methods were improved and widely replaced phenetics in the field of evolutionary taxonomy [36]. Some analyses from numerical taxonomy, such as neighbor joining analyses were subsequently adopted by cladistics for simple and fast tree analyses that avoid computational costs. However, inconsistencies of distance measurements in phylogenetic tree reconstructions have been pointed out. [37]

Molecular biology

Parallel to the methodological conflicts in taxonomy, molecular techniques started to improve, resulting in the first protein sequences presented by English biochemist Frederick Sanger in 1949 [38]. Following this success, attempts to compare sequences between species and potentially reconstruct species histories were made [39]. Around 1962-1965, Austrian biologist Emile Zuckerkandl and American chemist Linus Pauling discovered, via comparison of haemoglobin amino acid sequences, that these seem to change constantly with time [40]. Consequently, the idea developed that evolutionary histories of species could be dated using a differences in sequences as a molecular clock.

A dispute emerged between upcoming molecular biologists, that developed new techniques to infer evolutionary histories, and fathers of the evolutionary synthesis such as Mayr, Dobzhansky and Simpson. Consequently, arguments emerged regarding the levels at which selection acts, and Mayr would call for a separation of "functional" biology, asking "how?" and "evolutionary" biology asking "why?"[41]. This era has been retrospectively referred to as the "molecular war" [42]. However, although the reception of molecular evolution would not be completely positive at first, molecular phylogenetics was on its way to be established. In 1965, W. M. Fitch and E. Margoliash developed a method to reconstruct phylogenetic trees based on pairwise distances of protein sequences [43]. Subsequently in 1970, Fitch presented a parsimony approach for the reconstruction of phylogenetic trees [44]. This method constructs phylogenies according to the absolute minimum of sequence changes required to explain variation between species. By the time, DNA sequencing techniques were

proposed by Sanger [45, 46] and Maxam [47], Carl Woese, G. E. Fox and colleagues had worked on possibilities to use small ribosomal subunit for the inference of phylogenetic histories. In 1977, by the usage of rRNA oligonucleotide fingerprints Woese and Fox could identify a new kingdom of organism, the Archaebacteria [48]. A few years later, they would present the a tree reconstructed with the small ribosomal subunit [49]. With increasing computational power and the accumulation of protein and nucleotide sequences, more advanced methodologies would emerge. New techniques such as maximum likelihood [50, 51, 52] and Bayesian statistics [53] were introduced, aiming to refine the reconstruction of evolutionary histories and seek for the ultimate tree of life. In 2006, Cicarelli *et al.* presented a tree of life reconstructed from 31 concatenated orthologous genes found in 191 species of the three domains Archaea, Eukaryota and Bacteria. Genome sequences were and still are accumulating and it appears that variation in prokaryotes was vastly underestimated [54, 55]. The full scope of diversity in existing organism, cyanobacteria among them, still remains to be explored.

On Cyanobacteria

> *"I think it is now quite evident that the blue-green algae are not distinguishable from bacteria by any fundamental feature of their cellular organization"*
> **Roger Y. Stanier** (1970) [56]

The algal like appearance of many cyanobacterial species and the possession of chorophyll, a color-pigment also known from plants, resulted in their classification as "blue green algae" within the plant kingdom [57, 58]. A mistake that was finally resolved when Canadian microbiologist Roger Y. Stanier (1916-1982) wrote the sentence quoted above, in a letter to American botanist Peter Raven (1970). A few years later the bacterial nature of blue green algae should be established and indicated by their new name "Cyanobacteria".

A phylum with variable notations

During the first half of the 20th century and earlier, the nomenclature of modern cyanobacteria, conventionally named blue-green algae, was rather unstable. This is mainly caused by the unsteady classification of this phylum. Several division names have been used, most notably: Myxophyta, Schizophyceae, Chlorophyceae, Cyanophyta and Cyanobacteria. The endings *-phyta* or *-phyceae* indicated that blue green algae belong to the plant kingdom. The term Myxophyta is, according to Austrian botanist and cytologist, Lothar Geitler (1899-1990), the oldest version [57]. In the mid 19th century Myxophyta was replaced by Schizophyceae due to German biologist Ferdinand Cohn (1828-1889) [57, 59]. He suggested a close relationship between blue-green algae (Schizophyceae) and the — named by Swiss botanist Carl W. von Nägeli in 1857 — Schizomycetes (bacteria). Geitler on the other hand did not agree with Cohn's classification system and instead preferred the term Cyanophyceae, which indicated a closer relation to algal forms such as Chlorophyceae, Rhodophyceae and Phaeophyceae (green, red and brown algae) [57]. Since 1974, the notation cyanobacteria is the widely

accepted and used, though sometimes the out-dated name blue green algae may still appear in the literature. The term cyanobacteria as suggested Roger Y. Stanier's letter in 1970 (November 5, 1970, National Archives of Canada, MG 31, accession J35, vol. 6), and became officially established by "Bergey's Manual of Determinative Bacteriology" in 1974 [56, 60, 61].

The taxonomy of blue green algae

One of the first comprehensive works on cyanobacterial morphology was part of a plant systematics book series. From 1845 - 1871, Friedrich Traugott Kützing (1807-1892) published 19 volumes with about 1200 tablets in his "Tabulae Phycologicae", an elaborate study on the diversity of algae and other cryptogams [62]. Cryptogam is an obsolete taxonomic term for a non-monophyletic group of organisms that possess a hidden reproduction (no flower). Aside from several plant groups, cryptogams used to include fungi, blue green-algae and bacteria. The first two volumes of Kützings book series include detailed illustrations of some unicellular, but mainly multicellular cyanobacteria. Tablets 37-100 in volume 1 present cyanobacterial genera such as Spirulina, Oscillatoria, Phormidium, Lyngbya, Anabaena and Nodularia (Table 0.1). Tablets 1-84 of the second volume describe cyanobacterial genera, such as Nostoc, Rivularia and Stigonema (Figure 0.6 and Table 0.1).

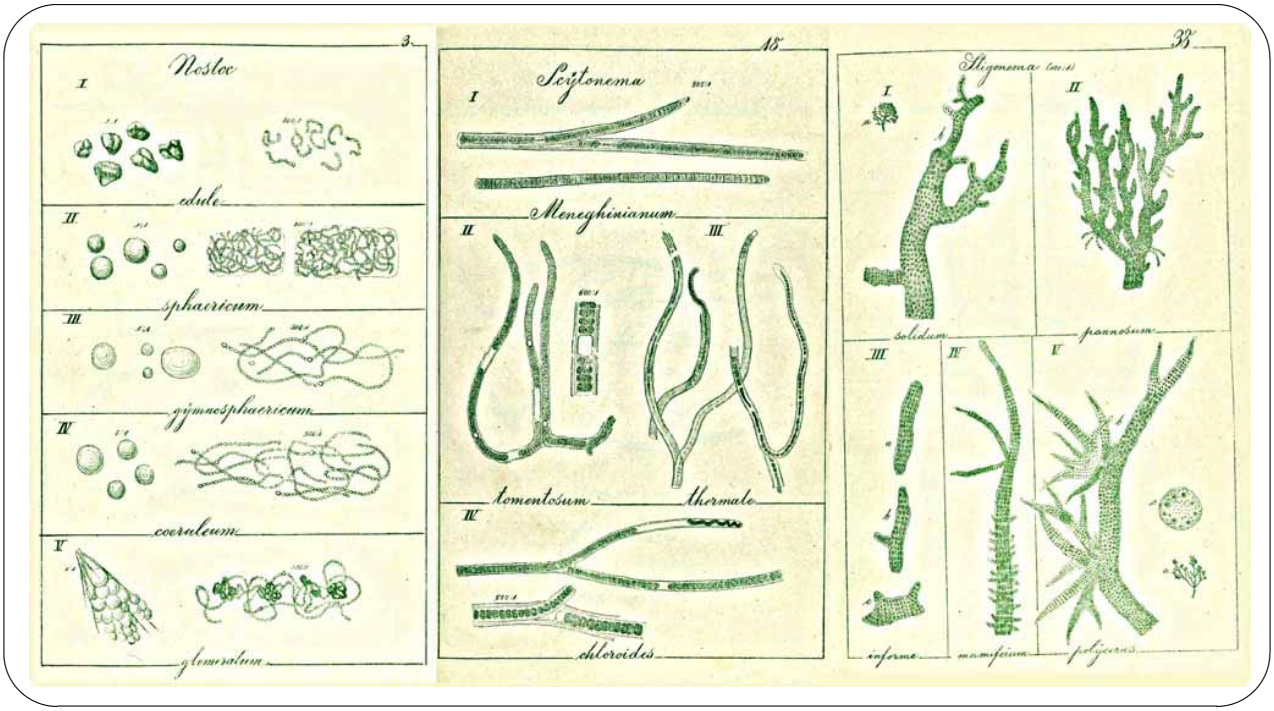


Fig. 0.6. Tablets showing cyanobacterial genera Nostoc, Scytonema and Stigonema adopted from **Kützing** (1850-1852)

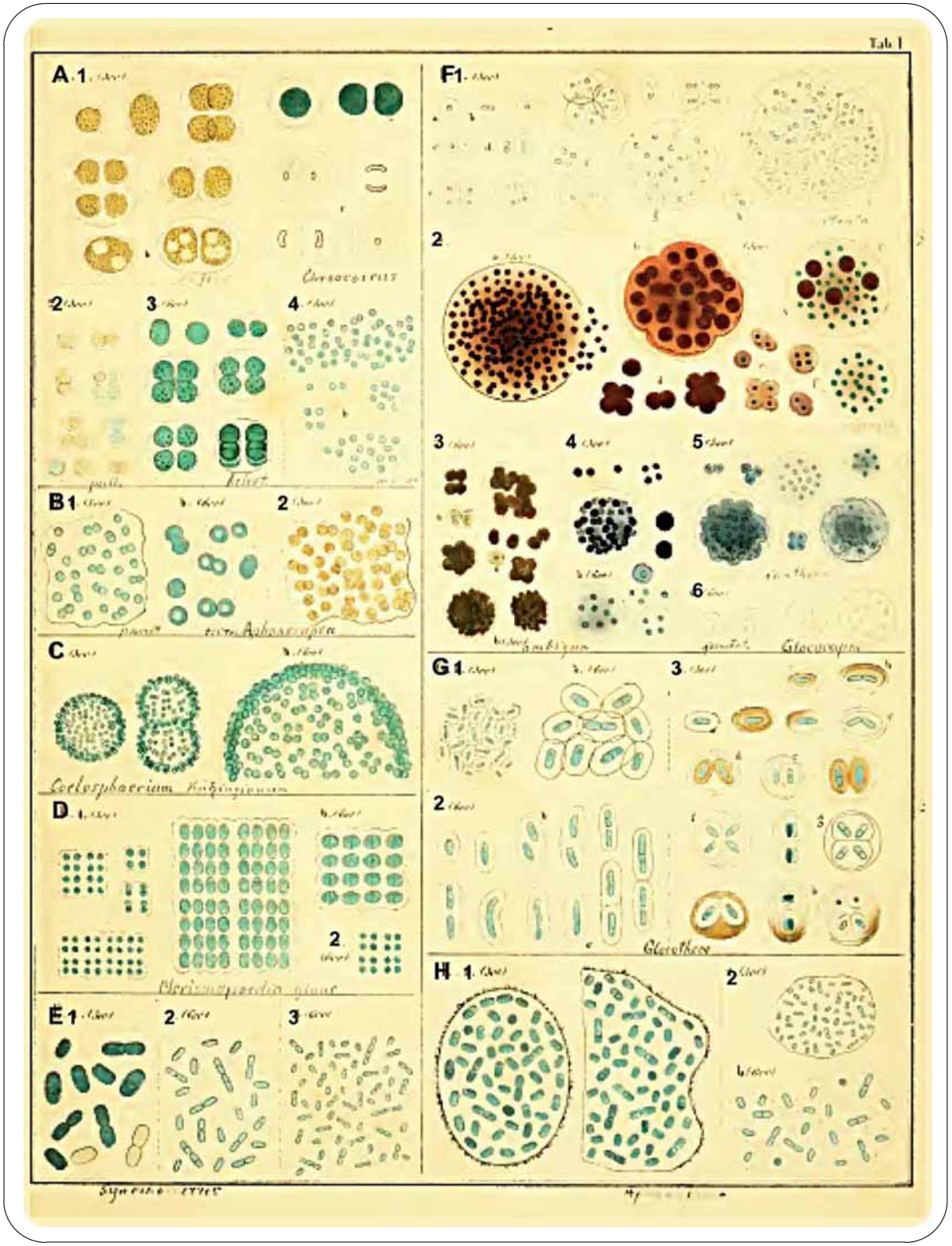


Fig. 0.7. Table 1. adopted from **Nägeli** (1849)

In 1849, Carl W. von Nägeli a Swiss botanist published a detailed systematic description on unicellular algae

[63]. The latter, as he claimed had been ignored by the systematic and physiological schools hitherto. He classified unicellular algae as single cells containing chlorophyll or a comparable color pigment, starch grains, and potentially color vesicles. Figure 0.7 displays Table 1 adopted from Nägeli's book *"Gattungen einzelliger Algen"*. The table contains several unicellular cyanobacteria, such as Chroococcus (Fig. 07-A) and Synechococcus (Fig. 06-E). The latter being genera whose names have been adopted from Nägeli's classification and are still in use today [64]. Another book series which did not concentrate on illustrations in forms of tablets as did Kützing but tried to inform about cytology, morphology and ecology of cryptogams was presented in two volumes (the second in three subvolumes) by Gottlob Ludwig Rabenhorst a German botanist (1806-1881), titled "Deutschlands Kryptogamen-Flora" (cryptogam flora of Germany) (1844-1848) [65, 66]. The original publication by Rabenhorst would be continued in several volumes after his death as *"L. Rabenhorst's Kryptogamen-Flora von Deutschland, Österreich und der Schweiz"*. The second volume *"Kryptogamen-Flora- Die Meeresalgen"* (1885) by F. Hauck [67] would include blue green algae (presented as order Schizophyceae, which contained two families, the Nostocaceae and the Chroococcaceae). The 1932 volume (14), of the *"Kryptogamenflora"*-series, written by the Austrian botanist and cytologist, Lothar Geitler (1899-1990) would designate cyanobacteria as "Cyanophyceae" indicating their perceived closer relation to other algae instead of bacteria [57]. The latter would become a popular reference for the nomenclature of cyanobacterial genera, termed the "Geitlerian system" [68]. According to "Bergey's Manual of Systematic Bacteriology" from 1984 (see page 16), the "Geitlerian system" would be followed by the short lived "Drouet system". A classification system of cyanobacteria which reduced the number of species and genera extensively [68]. In the 1970s, Roger Stanier and colleagues formulated a classification system that was based on the Geitlerian system, but would describe axenic cyanobacteria as part of the eubacteria in culture according to morphological, cytological, genetic, chemical, and physiological characteristics.

The phylogenetic relation of blue green algae

In the mid 19th century the phylogenetic position of cyanobacteria was still debated. Together with bacteria they were part of the plant kingdom, but whether they exhibited a closer relation to the "higher" plants or to the "more primitive" bacteria was not settled. Before Haeckels "Stammbaum der Organismen" from 1866 (Figure 0.3), living organisms would generally be separated into two divisions which had been classified by Linné as the kingdoms *Regnus Animalia* and *Regnus Vegetabilia* (animals and plants) [69]. Bacteria, which had already been discovered in 1676 by the Dutch tradesman and scientist Antonie van Leeuwenhoek (1632-1723), would be added to either of these kingdoms. Ernst Haeckel was the first who had classified bacteria separately from plants or animals. In his system of organism, bacteria, which he referred to as "Monera", were grouped together with other unicellular organisms such as "Protista" (Figure 0.3) [6]. However, he did not group cyanobacteria with the Monera, a common mistake made by many scientists [63, 57]. The German biologist Ferdinand Cohn (1828-1889) was one of the first who recognized the relation of the blue green algae and other chlorophyll lacking bacteria [70]. He suggested that Schizophyceae (blue green algae) together with Schizomycetes (bacteria) formed the Schizophyta [71, 59]. Originally, Ferdinand Cohn had separated bacteria

and blue-green algae into four groups: Spherobacteria (coccoid forms), Microbacteria (short, non-filamentous rods), Desmobacteria (long, filamentous rods), and Spirobacteria (spiral forms) [72, 59]. Just three years later he revised his classification, dividing Schizophyta in two distinct groups: Gloeogenae (single cells) and Nematogenae (filaments). Subsequently several scientists would recognize Cohn's suggestion. In 1900, Engler and Pratl confirmed this classification system of Schizophyta, which to their account marked the most primitive level within the plant kingdom [72]. In 1904, Ernst Haeckel would follow Cohn's suggestion — although not beyond doubt —, and group Schizophyceae within the Monera in his book "The wonders of life" [56, 73]. In 1910, E. Lemmermann separated Schizophyceae into two orders in his book "Kryptogamenflora" of Brandenburg: Coccogoneae, including modern Chroococcales, and Hormogoneae, including modern Oscillatoriales and Nostocales [74]. Further classification systems defined on the basis of Cohn's approach have been listed by R. E. Buchanan [75]. In his series published on "the nomenclature and classification of bacteria" Buchanan pointed out that classification systems of bacteria were still "chaotic". Buchanan agreed with the classification of blue green algae within the Schizophyta. Other scientists of that time, such as Hilda Hempl Heller, did not agree and demanded a clear separation of bacteria and Cyanophyceae [76]. Meanwhile, arguments on the relation of blue green algae would continue. The German botanist Ernst G. Pringsheim (1881-1940) argued for the possibility that bacteria and blue-green algae might have evolved via convergent evolution rather than sharing a common descent [77, 78]. Classification systems based solely on negative characters as presented by Roger Y. Stanier and the Dutch-American microbiologist C. B. van Niel (1897-1985), were according to Pringsheim not sufficient for the foundation of a taxonomic unit. Such negative characteristics found in bacteria and blue green algae are a lack of: a nucleus, sexual recombination and plastids. Stanier and van Niel emphasized the relationship of bacteria and blue-green algae which as they claimed had lost importance since its first recognition by Cohn [79]: "*It is at least certain that morphologically the Myxophyta resemble the true bacteria far more closely than they do any of the other algal groups*". Around the 1960s, technical advancements such as electron microscopy helped to classify cyanobacteria on a cellular level and identify them as belonging to the bacteria [61, 58].

Classification of prokaryotes

Until the 1940s, although the phylogenetic relation of of blue green algae was debated, most classification systems designated bacteria and Schizophyceae/Cyanophyceae to the plant kingdom. In 1938, H. F. Copeland would argue for a system where the Schizophyceae are part of a separate kingdom called Monera based on the absence of nuclei and on the assumption that they resemble descendants of *"whatever single form of life first appeared on earth"* [80]. An idea that, to his account, was based on previous work conducted by his father. Copelands suggestion was a four kingdom system of Monera, Protista, Plantae and Animalia.

Meanwhile between 1920-1940, hopes to find a valid phylogenetic classification for bacteria and consequently the ability to reconstruct the tree of life, were low [78]. In 1923-1925, David H. Bergey (1860-1937) had tried to develop a new classification system for bacteria based on structural and functional characteristics. A system

that to his account should replace useless earlier classification systems of the 19th century [68]. An approach that in the first half of the 20th century would receive various criticism. In 1949, Pringsheim criticised that classification methods used for plants would hardly work for bacteria [77]. A scepticism that was shared by other microbiologists such as Sergei N. Winogradsky (1856-1953) [81]. Winogradsky directed rather harsh criticisms against the classification approach applied by Bergey. In Winogradsky's opinion, "Bergey's Manual of Determinative Bacteriology" was no more than a *"delusion of a true phylogenetic ordering"* [56, 78]. Stanier and van Niel, opposed to that kind of "pessimism" directed towards bacterial phylogenetics, and held that indeed some relationships could be inferred even if bacteria show limited morphological characteristics [79]. They agreed with other microbiologists [81, 77] that "Bergey's Manual of Determinative Bacteriology" needed improvement. More precisely, they criticised the lack of important statements, on the *"absence of true nuclei"* which as they claim *"is perhaps the most important single morphological characteristic of these organisms"*, or on the *absence of sexual reproduction* [79]. Furthermore, they complained about confusing terminology, e.g. in the way the term "filament" is used. However, they recognized the increased usage of the manual for identification of bacteria at that time and the attempts by the editorial board to collaborate with bacterial specialists. Both indicators that led them to the assumption that Bergey's Manual might potentially become one of the most important classification literatures for bacteria [79].

In 1962, Stanier and van Niel rediscovered a classification system of organisms first mentioned by Èdouard Chatton (1883-1947), a French biologist, who distinguished two domains of organisms based on the presence of a real nucleus: prokaryotes and eukaryotes [82].

> *"It is now clear that among organisms there are two different organizational patterns of cells, which Chatton (1937) called, with singular prescience, the eukaryotic and prokaryotic type. The distinctive property of bacteria and blue-green algae is the prokaryotic nature of their cells. It is on this basis that they can be clearly segregated from all other protists (namely, other algae, protozoa, and fungi), which have eucaryotic cells."*
>
> **Stanier**(1962) [82]

The development of DNA sequence techniques in the 1970s [45, 47, 46] would pose new opportunities and encourage the reconstruction of phylogenetic histories of bacteria and to search for the universal tree of life [48, 49]. In 1977, Woese and Fox presented their results from rRNA oligonucleotide fingerprints that would lead to a reclassification of prokaryotes. Consequently, they would argue for three organismic "urkingdoms": Archaebacteria, Eubacteria and Eukaryota.

Bergey's Manual of Systematic Bacteriology

As suspected by Stanier and van Niel, Bergey's manual would gain importance in the microbiological community. The 8th edition of "Bergey's Manual of Determinative Bacteriology" established the term cyanobacteria [61, 83, 56]. The next edition of Bergey's Manual, was published in four separate volumes starting in 1984 [68]. With this new concept of splitting the manual into several volumes, the editorial committee intended to increase the descriptive quality. A notable change was the replacement of the name component 'Determinative Bacteriology' with the term 'Systematic Bacteriology', reflecting the notion of the publishers to assess

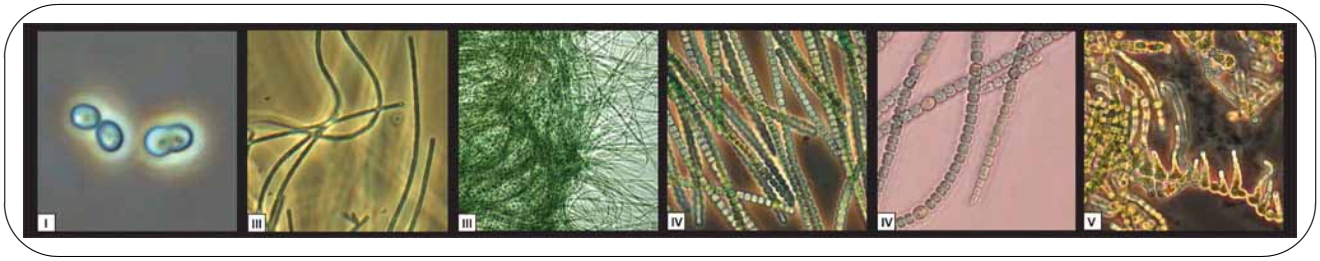


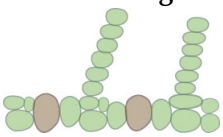
Fig. 0.8. Examples of morphological sections I, III, IV and V. Section III cyanobacteria were sampled by B.E. Schirrmeister from the North Sea.

evolutionary relationships of different prokaryotic groups [78, 68]. Cyanobacteria were presented in the third subvolume together with archaebacteria. Cyanobacteria were described as five subsections (Table 0.1;Figure 0.8), a classification system based on work presented by Rosmarie Rippka, Roger Stanier and colleagues in 1979 [84]. Section I and II comprise unicellular species while sections III to IV contain multicellular species. Furthermore, species from section IV and V are able to produce terminally differentiated cells called heterocysts. Species belonging to section V additionally show the ability to branch into multiple planes. The classification system was based on morphological, cytological, genetic, chemical and physiological characteristics [68]. Almost twenty years later, the second edition of Bergey's Manual of Systematic Bacteriology was published, again in four volumes with cyanobacteria and Archaea as part of the first volume [64]. The classification in this latest version of the manual is again divided in the five sections. Members of GROUP II - order Prochlorales have been integrated into subsections I and III in agreement with molecular phylogenetic results which identified members of the Prochlorales within the cyanobacteria, [85, 86]. Originally this group had been separated from all other cyanobacteria based on the occurrence of chlorophyll b and the lack of phycobilisomes (cyanobacterial light harvesting antenna) [86].

The international code of nomenclature for cyanobacteria

For a long period, Linne's binominal nomenclature system was used exclusively. Around 1867, rules for the notation of plants were established by Swiss botanist Alphonse de Candolle and introduced on a meeting of the International Botanical Congress [87]. After years of debate a real International Botanic Code was presented in 1905. The current *International Code of Botanic Nomenclature* (ICBN) derives mainly from a version from 1930 (Cambridge Code). The latest version was adopted by the 17th International Botanical Congress in Vienna, Austria, in 2005 (Vienna Code) [88]. Originally all bacteria would be named according to the botanical code [89]. However, the botanical code relies mainly on morphological descriptions. For the typification of species, holotypes should be conserved in a herbarium or an illustration. Under the Rules and Recommendations for Typification in the Vienna Code it says: *"Type specimens of names of taxa must be preserved permanently and may not be living plants or cultures."* (Division II, chapter II, section 2, Article 8.4) [88]. To simplify classification of bacteria in culture, microbiologists recognized an *International Code of Nomenclature of Bacteria and Viruses*, in 1958 [89]. Nevertheless, this code did not apply for cyanobacteria. Classification of the latter continued on basis of the botanical code which consequently, resulted in difficulties for the description of cyanobacterial

Tab. 0.1. Classification of cyanobacteria adopted from Bergey's Manual of Systematic Bacteriology 1989. Shown are only orders and genera (respective families of genera are not listed)

GROUP I		
I - Order Chroococcales 	**II - Order Pleurocapsales** 	**III - Order Oscillatoriales** 
unicellular	*unicellular, multiple fission*	*multicellular*
Chaemosiphon Gloeobacter Synechococcus group Gloeothece Cyanothece group Gloeocapsa group Synechocystis group	Dermocarpa Xenococcus Dermocarpella Myxosarcina Chrooccidiopsis Pleurocapsa group	Spirulina Arthrospira Oscillatoria Lyngbya Pseudanabaena Starria Crinalium Microcoleus
GROUP I		GROUP II
IV - Order Nostocales 	**V - Order Stigonematales** 	**Order Prochlorales**
multicellular, terminal differentiation	*multicellular, terminal differentiation, branching*	
Anabaena Aphanizomenon Nodularia Cylindrospermum Nostoc Scytonemataceae Rivulariaceae	Chlorogloeopsis Fischerella Stigonema Geitleria	Prochloron Prochlorothrix

species in culture, as Rippka *et al.* pointed out [84]. An attempt to classify cyanobacterial species under the rules of the *International Code of Nomenclature of Bacteria* (ICNB) by Stanier in 1978 [90] was rejected by many phycologists such as Geitler, Gobulic and Lewin [91, 68]. Potential problems that would arise from switching the nomenclature code were, amongst others the loss of previously defined taxon-names. Therefore in 1982, it was suggested that both nomenclatures should be applied for cyanobacterial classification [92]. Consequently, cyanobacterial strain/species typification would have been conducted on the basis of either ICBN or ICNB code using cultures or permanently preserved samples [93]. But, the call for a clear classification system for cyanobacteria is ongoing with argumentations for either code [94, 93]. The result is a partly chaotic cyanobacterial taxonomy. Improvements of this system could encourage scientific efforts to identify new species and to expand the knowledge on the true cyanobacterial diversity.

Cyanobacterial ancestry

For a long time, the existence of chlorophyll in blue-green algae was taken as evidence for a close relation to plants instead of bacteria [63, 57]. Doubts regarding this classification had accumulated [80, 79, 82], when in 1967 a suitable explanation for the occurrence of chlorophyll in cyanobacterial cells was described by American biologist Lynn Margulis (at that time named Lynn Sagan) (1938-2011) [95]. Margulis had recovered the idea first formulated in 1905 by the Russian botanist Konstantin Mereschcowsky (1855-1921), which suggested an origin of eukaryotic organells from bacteria via "symbiogenesis" [95, 96, 97, 56, 2]. In case of cyanobacteria this "Endosymbiosis Theory", described the origin of eukaryotic chloroplasts through engulfment of cyanobacteria via endosymbiosis [95, 98, 99]. Barghoorn and Tyler had described microfossils which seem to depict remnants of cyanobacterial relatives and supposedly date back more than 2.0 billion years [100]. The occurrence of such ancient blue green algae fossils in combination with geochemical evidence led to the conclusion that atmospheric oxygen accumulated around 2.1-1.7 billion years ago, probably earlier [101]. Sagan hypothesized that the accumulation of oxygen through cyanobacterial photosynthesis led to the evolution of "*different types of microbes...(including protomicrobes and prokaryotic aerobic algae)*". Sagan presented cytological evidence that led her to the conclusion that: "*Blue-green algae themselves maybe considered free-living prokaryote counterparts of plastids*" [95]. Although no fossil missing link for the endosymbiosis event was found, further studies on what appeared to be cyanobacterial fossils followed and dated the existence of cyanobacteria partly back to the Archean Eon, prior 2.5 billion years [102, 103, 104, 105, 106, 107, 108]. More recent time estimations of the first distinct increase of atmospheric oxygen, assume that this "Great Oxidation Event" (GOE) occurred around 2.45-2.32 billion years ago [109, 110, 111]. These assumptions inferred on the basis of geological and geochemical evidence, point to an origin of cyanobacteria prior to 2.0 billion years ago.

Molecular phylogenetics of cyanobacteria

In 1988, Giovannoni *et al.* conducted one of the first comprehensive cyanobacterial phylogenetic studies [112]. By that time eight cyanobacterial 16S rRNA sequences were known. Giovannoni *et al.* sequenced further cyanobacterial species to increase the presented evolutionary diversity of the phylum. Their inferred phylogenetic tree showed that morphological characteristics at least for sections I and III were phylogenetically non-informative. Furthermore, *Gloeobacter violaceus*, a unicellular cyanobacterium that lacks thylacoid membranes [113] grouped close to a eubacterial outgroup. Giovannoni suggested that the photosystem of Gloeobacter may resemble characteristics of a cyanobacterial most recent common ancestor and that living cyanobacteria are the result of an ancient adaptive radiation. His phylogenetic results were confirmed by phylogenetic studies that followed [114, 115, 85, 116]. Various analyses tried to identify truly monophyletic cyanobacterial groups using more advanced methods which incorporated evolutionary sequence patterns, such as multiple substitutions of nucleotides [117, 96, 85, 115]. In 1999, Honda *et al.* used neighbour joining and maximum likelihood analyses to infer phylogenetic relationships of 45 cyanobacterial species on the basis of 16S rRNA sequences [115]. The polyphyly of the subsections I and III was supported by their analyses. Furthermore, the phylogenetic tree indicated that cyanobacterial genera such as Synechococcus and Oscillatoria did not rep-

resent phylogenetically supported genera (non monophyletic). Again *Gloeobacter violaceus* was grouped at a position close to the eubacterial outgroup. The same year, phylogenetic studies presented by Turner *et al.* [85] confirmed once again the basal grouping of *Gloeobacter violaceus* and introduced new unicellular sister species of *Gloeobacter* which had been isolated from hot springs. Again, morphological sections I and III were spread across the phylogenetic tree. After publications of Ishida *et al.* in 2001, and an article by Gugger *et al.* titled "Polyphyly of true branching cyanobacteria (Stigonematales)" in 2004, monophyly of any of the cyanobacterial morphological sections was highly questionable.

The development of full genome sequencing techniques introduced the new field of comparative genomics. In 2004, the first cyanobacterial genomes were fully sequenced. Hess *et al.* conducted one of the first comparative genomic studies for marine pico-phytoplankton belonging to the genera Prochlorococcus and Synechococcus. Further phylogenetic and -genomic approaches followed, using increasing amounts of concatenated genes and improved reconstruction techniques [118, 119, 120, 121]. In 2005, Sànchez-Baracaldo *et al.* presented one of the first studies to estimate the evolution of morphological characteristics, such as salt tolerance, nitrogen fixation, and growth forms of cyanobacteria [118]. In a later study they refined their methodology and additionally estimated divergence times for cyanobacteria. Both analyses supported a unicellular most recent common ancestry from freshwater or terrestrial environments prior to 2.4 billion years ago [120]. In 2008, Swingley *et al.* presented a sophisticated comparative genomics study of 24 well studied cyanobacterial strains. The resultant phylogenomic tree was generally congruent to comparable 16S rRNA analyses. *Gloeobacter violaceus* and Synechococcus strains isolated from hot spring, again grouped close to a eubacterial outgroup. Marine picophytoplankton exhibited close relationships, as well as did terminally differentiated species.

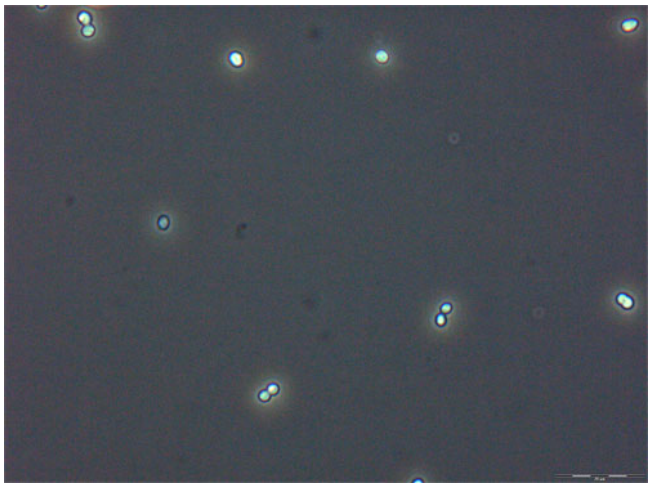
Where is cyanobacterial phylogenetics going?

Earth is inhabited by cyanobacterial taxa for more than half its existence. The evolution of oxygenic photosynthesis in this phylum, consequently led to the GOE, a key event in in the evolutionary history of life on the planet [110]. Furthermore, cyanobacteria contain a variety of morphotypes unique within the prokaryotes which consequently led to debates regarding their phylogenetic relation. Although, their prokaryotic character is today widely accepted [84, 68, 64], classification of cyanobacteria is still problematic [93]. Morphological classification systems help to characterize genera of this phylum, but are on their own insufficient to disentangle true evolutionary relationships. Results from molecular phylogenetics indicate that potentially none of the morphological subsections suggested by Rippka [84] are monophyletic. Although fossil data indicate that most of the morphological patterns already occurred 2.0 billion years ago [103, 122, 123, 108], evolutionary origins of these morphotypes are not understood (chapter II, chapter III). The development of molecular phylogenetics offers new possibilities to answer evolutionary questions. Recent advancements in genome sequencing have established the field of comparative genomics. However, cyanobacterial genome data are strongly biased. At present, the majority of sampled cyanobacterial species are unicellular strains belonging to section I. Data on the full diversity of the phylum cyanobacteria are still missing (chapter III). Therefore, results from full genome

comparisons enable a precise, but incomplete picture of cyanobacterial evolution. To investigate evolutionary questions in this phylum, to date, still relies on the taxonomic marker potential of 16S rRNA sequences (chapter I).

The following chapters, will hopefully enable the reader to grasp the importance and diversity of this phylum. I will elucidate (i) the association and capability of the small ribosomal subunit to and for the evolution of cyanobacterial morphotypes, reconstruct the early evolution of cyanobacterial morphotypes, and emphasize its consequences for life on this planet. For future phylogenetic or genomic studies, I recommend an elaborate sampling of cyanobacterial taxa. Genomic data on a cyanobacterial subset such as presented in this dissertation, covering the known diversity of this phylum, would clarify evolutionary relationships of cyanobacteria and contribute substantially to an understanding of the early evolution of life at the Archean/Proterozoic boundary.

CHAPTER I



Gene copy number variation and its significance in cyanobacterial phylogeny

Authors:

Bettina E. SCHIRRMEISTER, Daniel A. DALQUEN, Maria ANISIMOVA and Homayoun C. BAGHERI

(submitted)

Background: In eukaryotes, variation in gene copy numbers is often associated with diseases, but may also have positive effects. For prokaryotes, studies on gene copy number variation are rare. Previous studies have suggested that high numbers of rRNA gene copies can be advantageous in environments with changing resource availability, but further association of gene copies and phenotypic traits are not documented. We used one of the morphologically most diverse prokaryotic phyla to test whether numbers of gene copies are associated with levels of cell differentiation.

Results: We implemented a search algorithm that identified 44 genes with highly conserved copies across 22 fully sequenced cyanobacterial taxa. For two very basic cyanobacterial species distinct phylogenetic positions were supported by identical protein coding gene copy numbers. Furthermore, increased ribosomal gene copy numbers showed a strong correlation to cyanobacterial morphotypes and 16S rRNA sequence copies exhibited extremely low variation within species. We compared our results for 16S rRNA to three other eubacterial phyla (Chroroflexi, Spirochaetes and Bacteroidetes). Based on Bayesian phylogenetic inference and the comparisons of genetic distances, we confirmed that cyanobacterial 16S rRNA paralogs and orthologs show significantly stronger conservation than found in other eubacterial phyla.

Conclusions: A higher number of ribosomal operons could potentially provide an advantage to terminally differentiated cyanobacteria. Furthermore, we suggest that 16S rRNA gene copies are homogenized by both concerted evolution and purifying selection. In addition, the small ribosomal subunit in cyanobacteria appears to evolve at extraordinary slow evolutionary rates, an observation that has been made previously for morphological characteristics of cyanobacteria.

Background

Many genes originated via gene duplication in both prokaryotes and eukaryotes. Evolution after gene duplication can follow several scenarios [124]. Subfunctionalization leads to gene copies evolving specialized functions, all of which are necessary for performing the original gene function. In the neofunctionalization scenario, one gene copy is preserved by purifying selection, while the other copy may evolve a novel function through rapid adaptation. Finally, a process known as pseudogenization acts to remove “unwanted" copies by accumulation of neutral mutations. Therefore, possible evolutionary fates for gene duplicates include gene conservation by purifying selection and gene divergence via relaxation of selective constraints (to evolve subfunction), positive selection (to evolve a novel function) or neutral evolution (to convert to a pseudogene)[124]. Conserved gene copies can be easily detected based on their high levels of sequence similarity, which typically occurs for genes whose products are needed in high concentrations. All gene copies are strongly expressed in such cases. Gene duplicates can maintain their identical function in two ways: by purifying selection which prevents the duplicates from diverging, alternatively through concerted evolution where frequent gene conversion maintains sequence identity within the genome. Gene copy number variants, have been frequently found and studied in humans [125], but are also known to exist in other eukaryotic organisms, such as mouse [126], maize [127], and yeast [128]. Studies on human copy number variants revealed that multiple gene copies are often associated with diseases [129, 130], but can also have positive effects as has been shown for salivary amylase genes [131]. Less is known about consequences of protein coding gene copy number variations in prokaryotes. Though there have been studies on variation of ribosomal RNA gene copy numbers and possible consequences [132, 133]. Bacteria exhibiting multiple rRNA gene copies seem to respond faster to resource availability [134]. Accelerated growth rate has been conjectured to be a result of high ribosomal copy numbers [135]. In *E.coli* it is known that more than one rRNA operon has to be functional to express sufficient ribosomes and achieve maximum growth. Bacteria generally possess fewer than 10 rRNA gene copies [136], though some *Proteobacteria* and *Firmicutes* may have as many as 15 copies of rRNA operons [133]. Furthermore, ribosomal RNA copy numbers have been suggested to be phylogentically informative [137].

Although potentially important effects of ribosomal copy numbers have been suggested in various studies, protein coding gene copies are less considered. This could be due to the assumption that selection for faster cell replication leads to genome reduction in prokaryotes [138], which would reduce the likelihood of survival of multiple gene copies. Indeed, a tendency towards genome reduction has been observed in endosymbiotic bacteria, and in free living prokaryotes including unicellular marine cyanobacteria [139]. However, contradictory conclusions have been made by Kou and colleagues [140]. They point out that neither could a correlation be detected between genome size and doubling time, nor do endosymbionts with small genome size benefit from short cell division times. Instead, a lack of large prokaryotic genomes could be the result of selection acting on an upper limit of genome size. Thus, if there is no selective genome reduction in prokaryotes, multiple gene copies might be more widely distributed and of greater importance for prokaryotes than is believed so far. Full genome data are accumulating, and studies on copy number variation, as well as on its potential

consequences for prokaryotic phenotypes would add to our knowledge of bacterial evolution.

Among prokaryotes cyanobacteria depict one of the morphologically most diverse phyla. Several of their morphotypes seem to exist for over 2.00 billion years as indicated by a well preserved fossil record [103, 123]. Cyanobacteria inhabit diverse environments. Among others, they can be found in soil, marine or limnic waters ranging from arctic to tropic climate zones. They had (and still have) an exceptional influence on the planet due to their ability to conduct oxygenic photosynthesis and fix nitrogen. According to their morphology, cyanobacteria have been classified into five different "clades" [64], though molecular data indicate that probably none of the five clades is monophyletic [112, 85, 116, 141, 142, 143]. Therefore, we will refer to these clades hereafter as sections. Section I and II consist of unicellular cyanobacteria. Section II species can be distinguished from all other cyanobacteria based on their reproduction via multiple fission. Cyanobacteria belonging to section III to V exhibit filamentous growth. Across the five existing morphotype sections cyanobacteria exhibit several patterns of differentiation. The majority of extant cyanobacterial species control gene expression using a circadian clock. Additionally, several multicellular cyanobacteria developed mechanisms to differentiate not only temporarily, but also spatially. Trichodesmium is the only section III genus known, able to produce specialized cells ('diazocytes') in the middle of a filament [144, 145, 146]. Though, presence of these cells has been questioned elsewhere [147, 148]. In these reversibly differentiated cells, nitrogen can be fixed during the day. This process would not be possible otherwise, due to high oxygen sensitivity of the involved enzyme. The only form of terminal differentiation is observed in section IV and V cyanobacteria. In the absence of combined nitrogen, these organisms form heterocyst cells where nitrogen fixation can take place at any time. Given the morphological variety found in this phylum, we ask whether gene dosage (multiple gene copies per cell) might be associated with adaptive morphological strategies such as cell differentiation in cyanobacteria. Variation in 16S rRNA gene copy sequences and numbers has been reported previously for cyanobacterial genera [149, 150], but no phenotypic correlations were found. Little is known about protein coding gene copy numbers in cyanobacteria. In this study we searched for both ribosomal RNA and protein coding gene copy number variation in diverse species of cyanobacteria for which full genome sequences were available.

Ribosomal RNA gene copies were examined since it is known that they might occur in multiple copies and exhibit gene dosage effects [134, 136, 135]. Ribosomal RNA is an essential component of the ribosome, the protein manufacturing complex. Segments of genes within the rRNA operon are strongly conserved because of their functional relevance [151], whereas others show significantly higher substitution rates. The strong conservation makes aligning gene homologs from distant species easier and helps disentangle deep relationships. In contrast, faster evolving homologs can be studied only at the family or genus level. These unique features have made 16S rRNA gene sequences a favored taxonomic marker for prokaryotes [152]. Although rRNA sequence variation within a genome is low for most species [132], considerable intragenomic differences have been reported in some non-cyanobacterial species [153, 133]. This has led to the questioning of the reliability of 16S rRNA genes as a taxonomic marker. We examined sequence identity of rRNA genes within species

of cyanobacteria by conducting phylogenetic analyses and calculating phylogenetic distances. Results for cyanobacteria were compared to data from the prokaryotic phyla Chroroflexi, Spirochaetes, and Bacteroidetes. Paralogs of 16S rRNA genes are almost identical in cyanobacterial species and suggest a deviation from divergent evolution of gene copies. Investigating variation in copies of the internal transcribed spacer region (ITS), located between the 16S and 23S rRNA genes, suggests that both concerted evolution and purifying selection are viable hypotheses for the evolution of 16S rRNA in cyanobacteria. Furthermore, we observed an exceptionally strong sequence conservation in 16S rRNA orthologs within the cyanobacterial phylum. A level of conservation that could not be observed in any of the eubacterial phyla studied here.

Results and Discussion

Identification of conserved gene copies and their phylogenetic relevance

Aside from ribosomal genes, we identified 41 protein coding genes which possess multiple conserved gene copies in at least one cyanobacterial species (Additional File 1.1). From this total of 44 genes, only six showed significant correlations to morphological characteristics. Ribosomal RNA genes were the main class of genes exhibiting conserved gene copies that were significantly correlated to sections IV and V in cyanobacteria. Species capable of terminal cell differentiation conclusively exhibited four or five copies of ribosomal genes.

We investigated conserved gene copies that exhibited $\geq$ 90%(not shown), $\geq$ 95%(not shown) and $\geq$ 98% amino acid sequence identity within a genome. Results varied mainly in numbers of transposase gene copies detected. Therefore, results of gene copies with an identity of $\geq$ 98% within a genome and $\geq$ 50% between species are presented here. For these genes, we mapped copy numbers in relation to the phylogenetic position within cyanobacteria (Figure 1.1). The highest number of gene copies (24) was found for a transposase encoding gene in *Microcystis aeruginosa*. Transposases are enzymes that catalyze the movement of transposable elements. Previous studies have estimated that genes encoding for transposases are the most widespread genes, and often occur as multiple copies [154]. Almost half of the conserved gene copies identified in this study were transposase encoding genes. The frequency of transposase genes varied between different species. *Microcystis aeruginosa* possessed various transposase genes, whereas strains belonging to the genera Synechococcus and Prochlorococcus, and *Cyanobacterium sp. UCYN-A* seem to exhibited fewer transposase gene copies.

Synechococcus sp. JA-3-3Ab, a unicellular cyanobacterium isolated from a hot spring in Yellow Stone National Park [155, 156], exhibited a pattern of gene copy numbers that generally deviated from the pattern observed in other Synechococci. It shared identical copy numbers of protein coding genes with *Gloeobacter violaceus*. These included a series of not yet annotated genes missing in all other cyanobacteria. This pattern of almost identical conserved gene copy numbers supports other phylogenetic studies that place these two species close to each other at the base of the cyanobacterial phylogenetic tree [119, 157, 158]. Phylogenomic studies point to the same result [119]. In addition, using 16S rRNA sequences we observed a close phylogenetic relationship of *Gloeobacter violaceus* and another isolate from a hot spring in Yellowstone National Park, *Synechococcus*

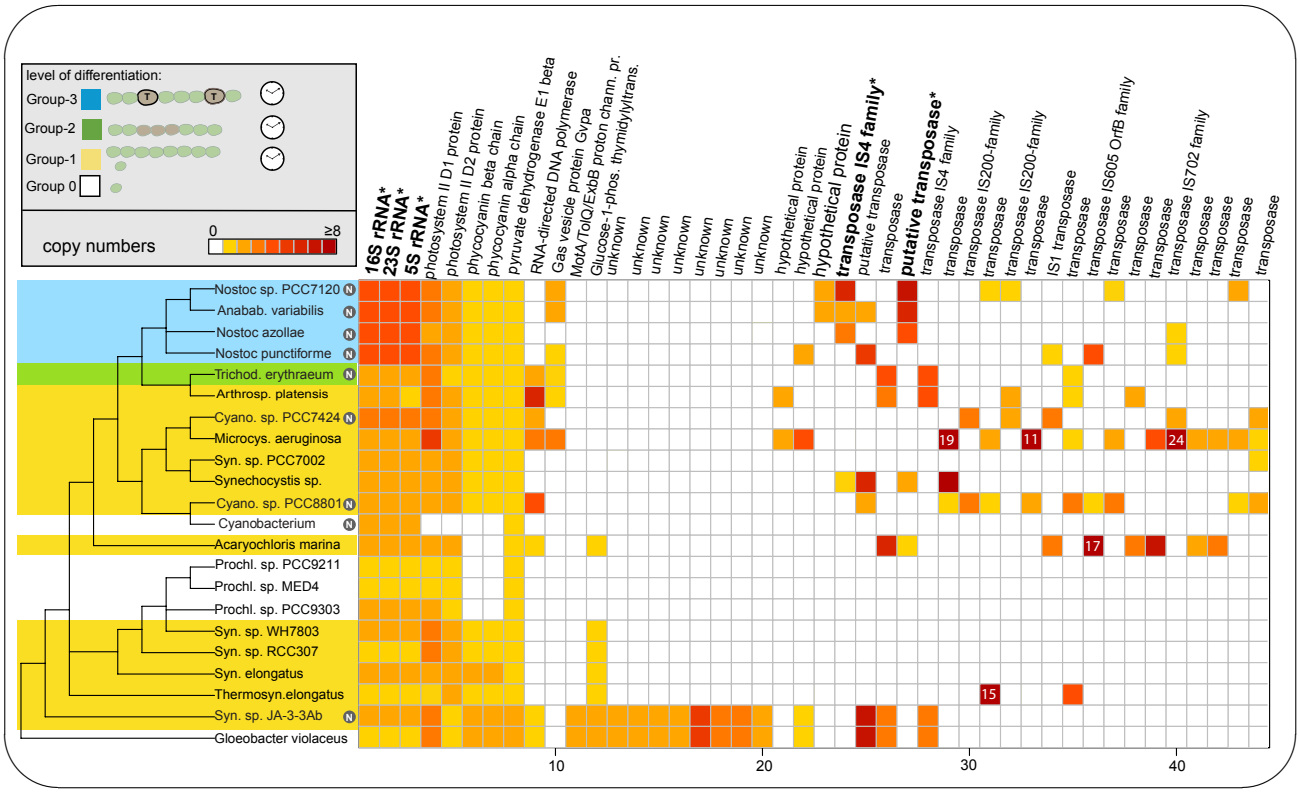


Fig. 1.1. Conserved paralogs in cyanobacteria. Distribution of gene copy numbers within and across cyanobacterial genomes. On the left side cyanobacterial cladogram is shown, emphasizing the different morphological groups. Species of group G1 exhibiting circadian rhythm are displayed in a yellow box. *Trichodesmium* exhibiting reversible differentiation is shown in a green box (group G2) and cyanobacteria of group G3 which are able to terminally differentiate, are displayed in a blue box. The letter 'N' marks species capable of nitrogen fixation. Conserved copy numbers of genes are shown in a color plot ranging from yellow indicating a single gene to dark red denoting 8 copies or more. In cases where gene copy numbers exceed 8, values are given in white letters. Corresponding species names are written on the left and gene names are written on top. Genes in bold and marked by a "*" are positively correlated to terminal differentiation.

sp. P1 [159]. The phylogenetic distance of *Gloeobacter violaceus* to other extant cyanobacteria has been pointed out before [160]. Major differences involve the light harvesting machinery. *Gloebacter violaceus* lacks thylacoid membranes [113], and various genes from photosystems I and II.

Several genomes with more than one ribosomal gene copies were identified. Cyanobacterial taxa used in this study exhibited one to four conserved rRNA gene copies (1.1, Table 1.1). Position of ribosomal gene copy numbers across the Bayesian tree (Figures 1.1 and 1.2) were phylogenetically non-informative. However, four rRNA copies could only be observed in terminally differentiated species. Additional data on 16S rRNA copy numbers shown in the rrn-database, confirmed these findings and furthermore reported five copies for several cyanobacterial species belonging to sections IV and V. Aside from 16S rRNA data, no further information was obtained, because these taxa have not been fully sequenced, yet [161].

Correlation of copy numbers to terminal differentiation

To test whether ribosomal RNA or protein coding gene copy numbers can be associated to species capable of terminal differentiation, we visualized the distribution of ribosomal gene copy numbers and tested for possi-

Tab. 1.1. Data of cyanobacterial 16S rRNA gene sequences d^1: Largest distance between gene copies within the genome. F: Coordinates for the 16S rRNA genes on the forward strand of the chromosome. R: Coordinates for the 16S rRNA genes on the reverse strand of the chromosome.

Species	Group	Genome size	Nr. of copies	d^1	F	F	R	R	Accession nr.
Acharyochloris marina MBIC11017	G1	8.36	2	0	5,636,175		1,409,149		CP000828.1
Anabaena variabilis ATCC 29413	G3	7.10	4	0	1,002,918	3,894,075	2,808,379	5,435,874	CP000117.1
Arthrospira platensis NIES 39	G1	6.80	2	0			2,584,861	3,509,612	AP011615
Cyanothece sp. PCC 7424	G1	6.52	3	0.001	1,328,842	3,465,297	2,494,023		CP001291.1
Cyanothece sp. PCC 8801	G1	4.81	2	0	3,806,018		2,484,826		CP001287.1
Gloeobacter violaceus PCC 7421	G0	4.70	1				1,571,231		BA000045.2
Microcystis aeruginosa NIES-843	G1	5.80	2	0.003	1,885,807		3,597,272		AP009552.1
Nostoc azollae 0708	G3	5.53	4	0	830,919	2,217,271	979,079	2,979,417	CP002059.1
Nostoc punctiforme PCC 73102	G3	9.01	4	0.002	2,021,489	6,085,170	5,515,629	6,502,973	CP001037.1
Nostoc sp. PCC 7120	G3	7.20	4	0	2,375,734	2,500,525	4,918,283	5,945,700	BA000019.2
Prochlorococcus marinus MIT 9211	G0	1.70	1		342,283				CP000878.1
Prochlorococcus marinus MIT 9303	G0	2.70	2	0	243,682		1,938,786		CP000554.1
P. marinus subsp. pastoris str. CCMP1986 (MED)	G0	1.70	1		313,061				BX548174.1
Synechococcus elongatus PCC 6301	G1	2.70	2	0	1,656,455		1,050,801		AP008231.1
Synechococcus sp. JA-3-3Ab	G1	2.90	2	0	2,310,397		1,110,127		CP000239.1
Synechococcus sp. PCC 7002	G1	3.40	2	0	1,461,361		2,909,371		CP000951.1
Synechococcus sp. RCC307	G1	2.20	1		348,765				CT978603.1
Synechococcus sp. WH 7803	G1	2.40	2	0	534,563		2,019,450		CT971583.1
Synechocystis sp. PCC 6803	G1	3.97	2	0	3,325,053		245,2187		BA000022.2
Thermosynechococcus elongatus BP-1	G1	2.59	1				2,335,243		BA000039.2
Trichodesmium erythraeum IMS101	G2	7.80	2	0	3,137,164		4,601,878		CP000393.1
Cyanobacterium UCYN-A	G0	1.40	2	0	638,681		3,507		CP001842.1

ble correlations to morphotypes (Figure 1.3). We calculated potential correlations of gene copy numbers and morphotypes for all genes identified in this study. We divided cyanobacteria into four morphological groups according to their mode of differentiation. Group 0 (G0) exhibits no mode of differentiation and contains solely unicellular species. Group 1 (G1) consists of species from section I to III which control gene expression via a circadian rhythm, but lack any other form of differentiation. Group 2 (G2) is formed exclusively by the genus *Trichodesmium* which is able to form temporarily differentiated cells for nitrogen fixation. The last group (G3) contains species from section IV and V which are able to produce terminally differentiated cells.

Using Spearman's rank correlation coefficient (ρ) and Pearson's correlation coefficient (R), we estimated a potential correlation of copy numbers to the defined morphological groups. Both tests indicated significant correlations to morphological groups for all ribosomal genes and two transposase coding genes. Furthermore, Spearman's ρ attested a significant correlation to morphology for photosystem II reaction center D2 protein (ρ=0.62), and a weaker correlation to Gas vesicle protein GVPa (ρ=0.58) coding genes. A significant Pearson's correlation was found for a gene coding for a hypothetical protein (R=0.58). In Figure 1.3 distributions of ribosomal RNA gene copy numbers across morphological groups are presented as boxplot graphics with correlation coefficients, and p-values shown. All taxa capable of terminal differentiation exhibited four copies of ribosomal RNA genes. Correlation coefficients for 16S and 23S rRNA genes were ρ=0.74/R=0.86, in both cases, and ρ=0.63/R=0.8 for the 5S rRNA genes. Including additional data from the rrn-database [161] (Additional File 1.2), resulted in an even stronger correlation of 16S rRNA gene copy numbers to cyanobacterial species capable of terminal differentiation (ρ=0.87/R=0.9; Additional File 1.3). Cyanobacteria belonging to section IV and V form terminally differentiated cells (called heterocysts) in the absence of fixed nitrogen. In these cells

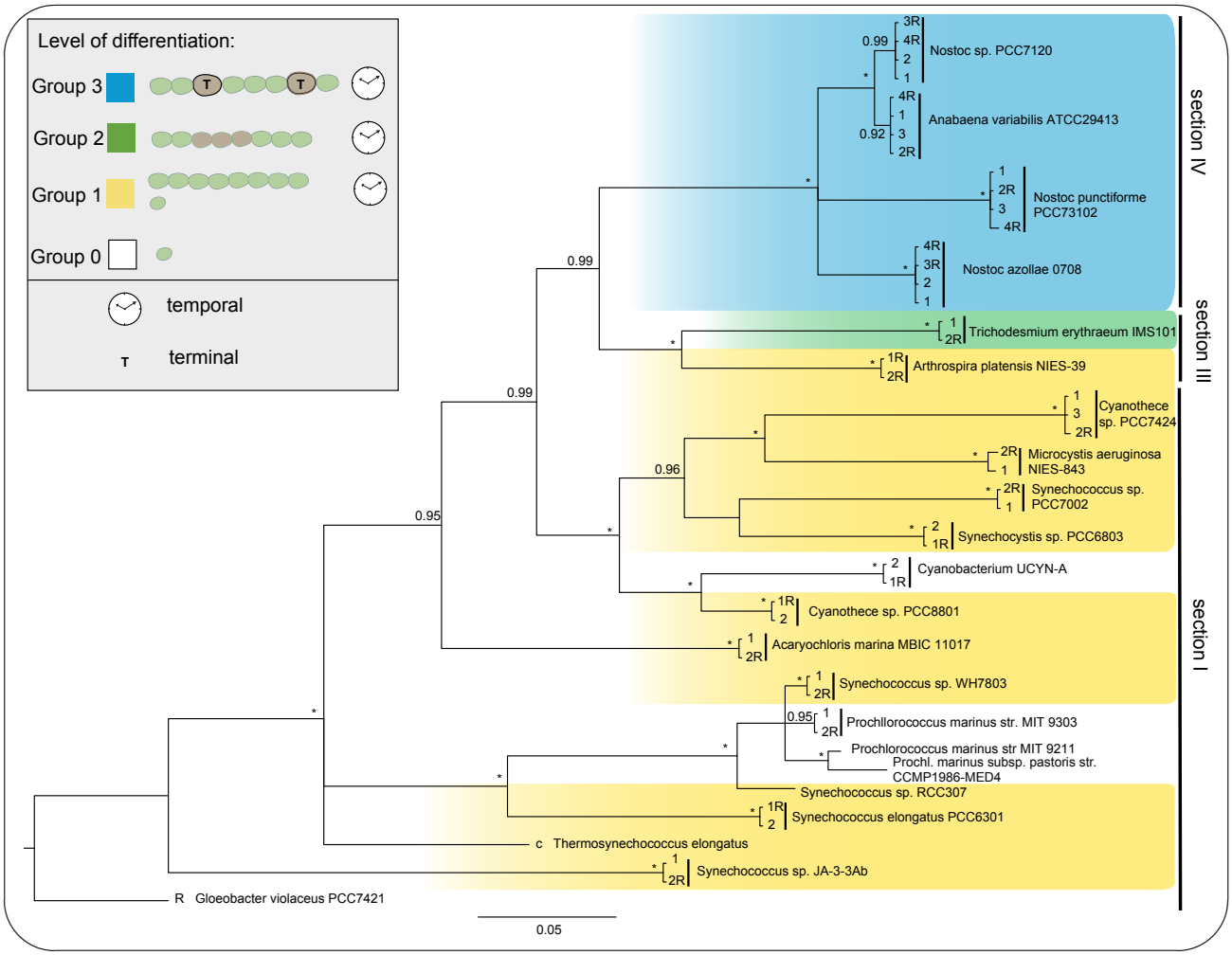


Fig. 1.2. Cyanobacterial tree including all 16S rRNA gene copies. Cyanobacterial tree including all 16S rRNA copies, reconstructed using Bayesian analysis. Posterior probabilities > 0.90 are displayed on the nodes. Colors indicate species-groups according to differentiation level. Species in yellow boxes control gene expression only via a circadian rhythm. Genus *Trichodesmium* shown in a green box is able to produce temporarily differentiated cells, called 'diacocytes'. Multicellular species able to form terminally differentiated cells are shown in blue boxes. The letter "R" denotes gene copies that are positioned on the reverse DNA strand. Multicellular, terminally differentiated cyanobacteria are the only species exhibiting four copy numbers. Regardless of morphology, 16S rRNA sequences are highly conserved within each genome.

oxygen sensitive nitrogen fixation can take place while neighbouring cells conduct oxygenic photosynthesis. These heterocystous cells undergo various structural and physiological alterations to protect nitrogenase from oxygen in a 'microanaerobic' environment. As a result they lose their ability to conduct photosynthesis and to divide. Multiple rRNA gene copies could have positive effects during heterocyst formation, the same way as they help *E.coli* to achieve maximum growth [135], and increases responses to changing environmental conditions [134]. An increased amount of functional ribosomal operons likely depicts an advantage in the process of cell differentiation, during which expression of various genes is upregulated [162].

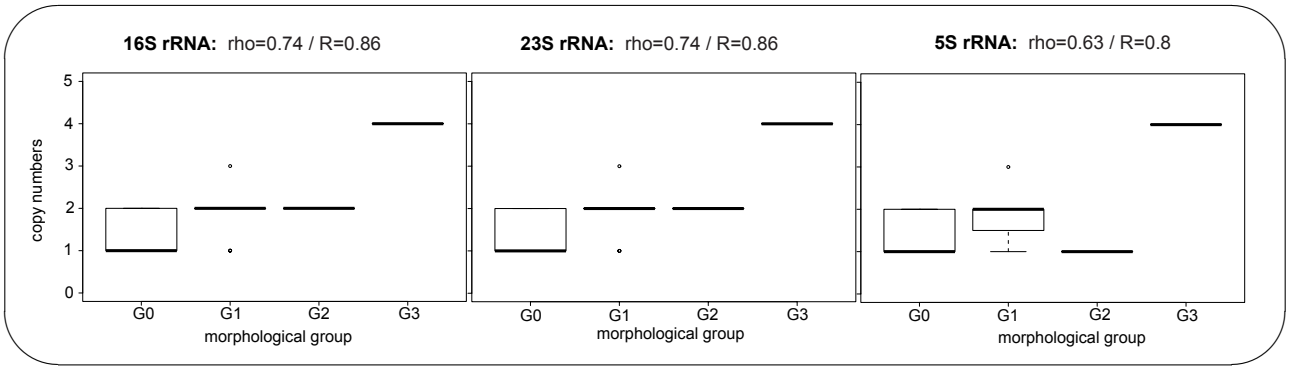


Fig. 1.3. Dispersion of gene copy numbers in different groups of differentiation. A boxplot representation of the gene copy number dispersion across the previously defined morphological groups. Shown are dispersions for genes from the rRNA operon. Spearman's rank correlation coefficient (ρ) and Pearson's correlation coefficient (R) are displayed above the corresponding graph. Positive correlation coefficients of rRNA gene copies to terminally differentiated cyanobacteria are supported.

Strong conservation of 16S rRNA copies

To identify whether the strong intraspecific rRNA conservation is exclusive to cyanobacteria, we investigated 16S rRNA sequences of cyanobacteria in more detail and compared our results to other eubacterial phyla. Investigated cyanobacteria, exhibit one to four 16S rRNA copies per genome. Using Bayesian analyses we reconstructed phylogenetic trees to evaluate the divergence of 16S rRNA gene copies within and between cyanobacterial taxa. The inferred Bayesian consensus tree is displayed in Figure 1.2. Unicellular species partition in two major groups: species belonging to the marine pico-phytoplankton genera *Synechococcus* and *Prochlorococcus*, and members of the genera *Synechocystis*, *Cyanothece* and *Microcystis* which show a closer relation to multicellular cyanobacteria. All multicellular species studied here are closely related, and species capable of terminal differentiation form a monophyletic group. Comparisons of our study to previous findings show high similarities. Our results agree with a comparative phylogenomics approach used by Swingley *et al.* [119], a consensus tree of concatenated sequences presented by Blank and Sànchez-Baracaldo [120], and are highly similar to 16S rRNA analyses conducted by Schirrmeister *et al.* [159]. Using a larger taxon set [159], we inferred polyphyletic groupings of undifferentiated multicellular species belonging to section III. This however is not deducible from the taxonomically more limited full genome data set used in the present study.

In cyanobacteria 16S rRNA sequences were highly conserved within a genome. Three species showed minor nucleotide differences. The two 16S rRNA copies of *Microcystis aeruginosa* differed by four 'single nucleotide polymorphisms' (SNPs), in *Cyanothece sp.* PCC 7424 one SNP was detected, and in *Nostoc punctiforme* one 16S copy possessed two SNPs. The differences are visualized in a molecular distance matrix in Figure 1.4. 16S rRNA copies within species were identical for the majority of taxa (shown in yellow) and can be clearly distinguished from gene copies belonging to different species. Furthermore, using the whole dataset we calculated mean distances within strains (d_W) and between strains (d_B). Results are presented in Table 1.2. Significance of differences in sequence distances found within and between cyanobacterial strains were estimated using bootstrap re-sampling of the original data set. Distributions of the resulting mean distances are displayed in

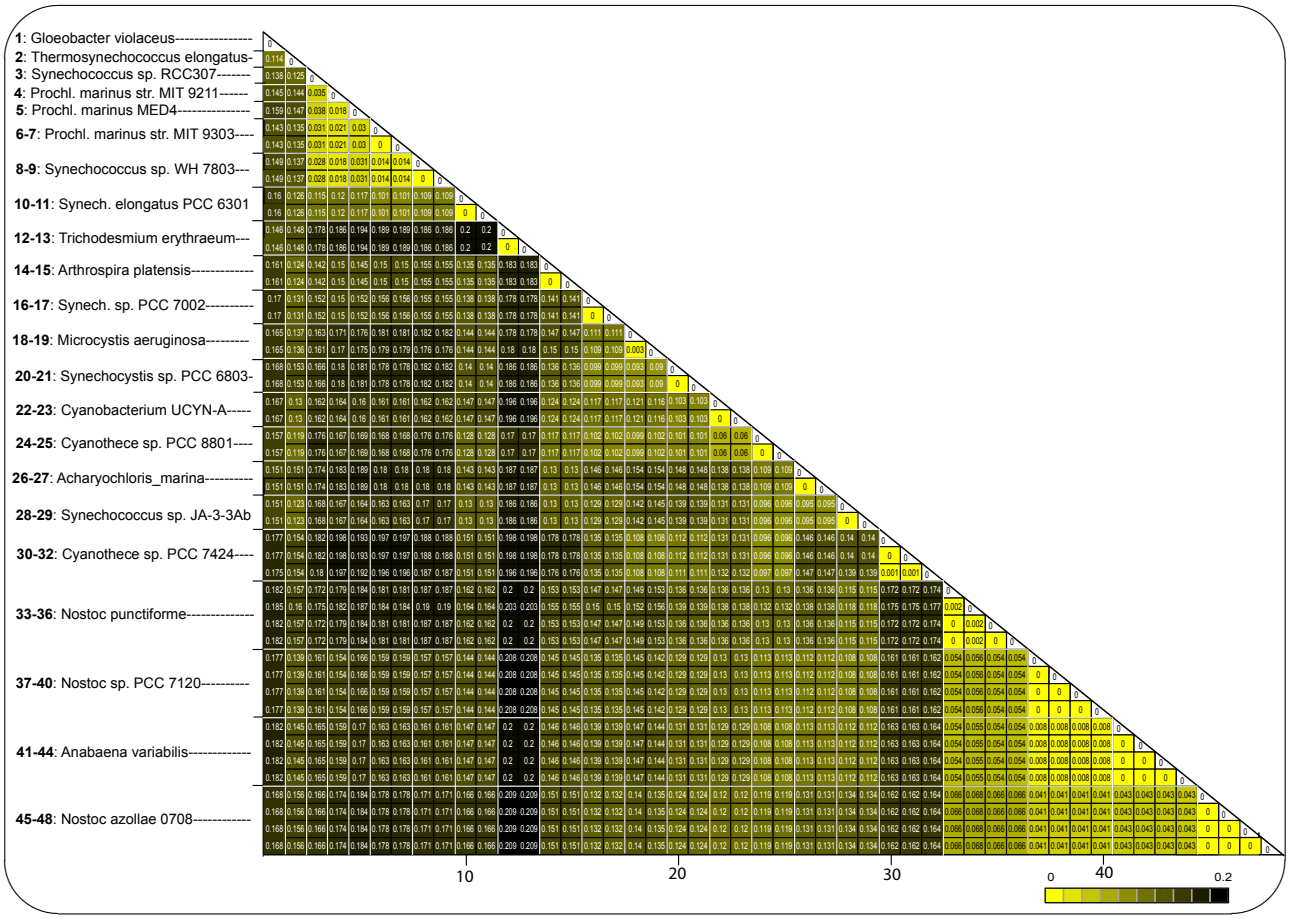
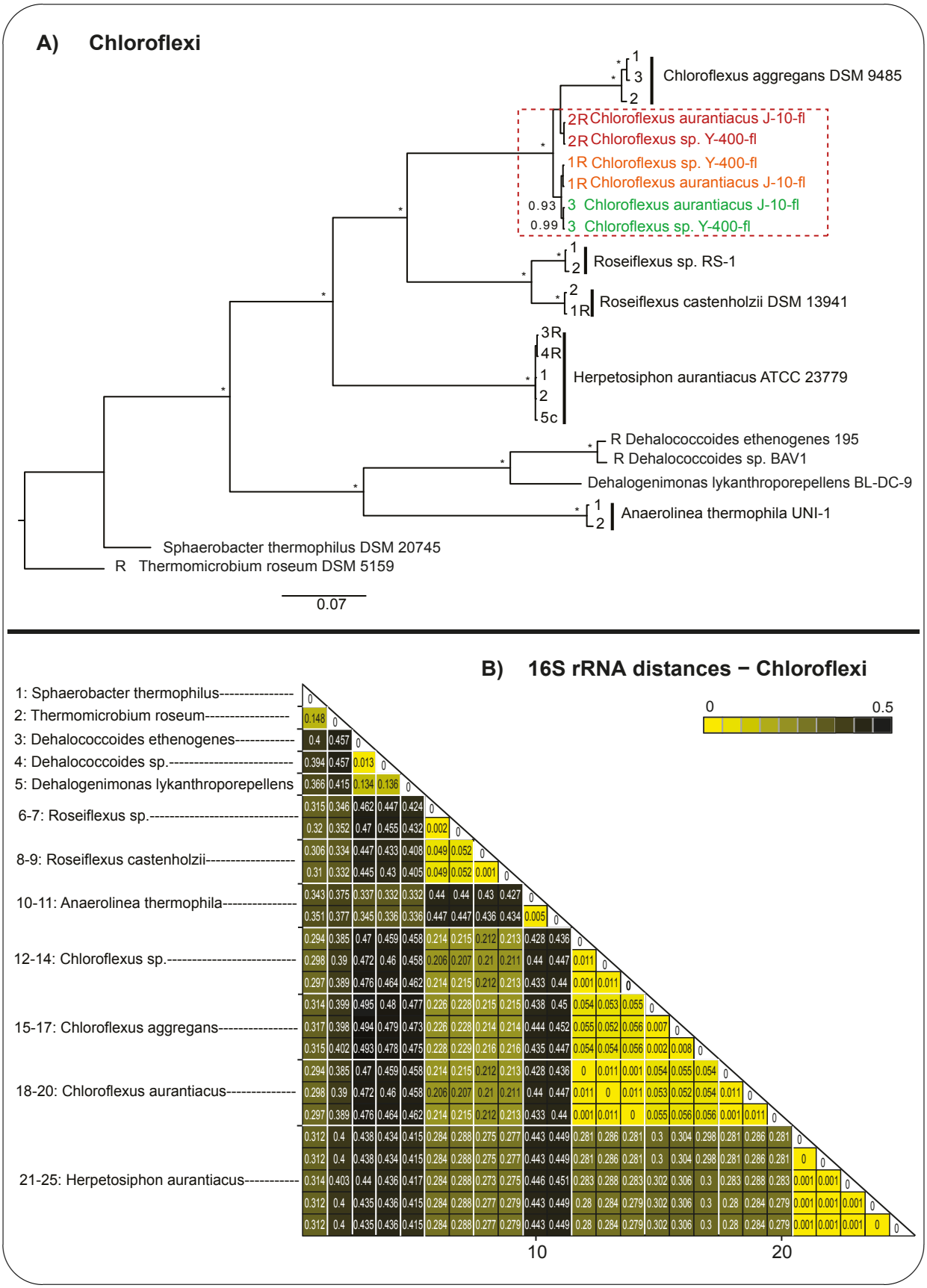


Fig. 1.4. Distance matrix of cyanobacterial 16S rRNA sequences. Distance matrix between 16S rRNA genes estimated based on K80 substitution model. 16S rRNA gene copy numbers range from one to four per cyanobacterial genomes studied. White lines separate sequence copies of different species. 16S rRNA sequences are highly conserved within species.

Additional Files 1.4 and 1.5. For each distribution, an overall mean distance was calculated (d_W^*,d_B^*). Mean distance of 16S rRNA sequences within species ($d_W = 0.0001$) is significantly smaller than between species ($d_B = 0.14$; Table 1.2). 95% confidence intervals of distributions obtained by re-samplings do not overlap. Although previous studies have claimed that variation within 16S rRNA sequences might affect reliability of this gene as a taxonomic marker [153, 133], this was not found for genera used in this study. Rather, the extreme sequence conservation of 16S rRNA gene copies from the same species supports 16S rRNA as a reliable genetic marker for the taxa analyzed here.

In order to verify the significance of our results for cyanobacteria, we compared phylogenetic and distance

Fig. 1.5 *(on the next page).* Phylogenetic tree and distance matrix of Chloroflexi including all 16S rRNA copies. (**A**) Phylogenetic tree of the eubacterial phylum Chloroflexi including all 16S rRNA copies, reconstructed using Bayesian analysis. On the nodes posterior probabilities > 0.90 are displayed. Colored taxa mark species where 16S rRNA copy numbers evolved rather via divergent evolution, than being homogenized within a strain via concerted evolution. The letter "R" denote gene copies that are positioned on the reverse DNA strand. (**B**) Distance matrix of Chloroflexi. Genetic distances have been estimated according to the K80 substitution model. White lines separate sequence copies of different species. 16S rRNA sequences are conserved within species, but exhibit more variation than found for cyanobacteria.


A) Chloroflexi
Chloroflexus aggregans DSM 9485
2R Chloroflexus aurantiacus J-10-fl
2R Chloroflexus sp. Y-400-fl
1R Chloroflexus sp. Y-400-fl
1R Chloroflexus aurantiacus J-10-fl
3 Chloroflexus aurantiacus J-10-fl
3 Chloroflexus sp. Y-400-fl
0.93
0.99
Roseiflexus sp. RS-1
Roseiflexus castenholzii DSM 13941
Herpetosiphon aurantiacus ATCC 23779
R Dehalococcoides ethenogenes 195
R Dehalococcoides sp. BAV1
Dehalogenimonas lykanthroporepellens BL-DC-9
Anaerolinea thermophila UNI-1
Sphaerobacter thermophilus DSM 20745
R Thermomicrobium roseum DSM 5159
0.07
B) 16S rRNA distances − Chloroflexi
0
0.5
1: Sphaerobacter thermophilus
2: Thermomicrobium roseum
3: Dehalococcoides ethenogenes
4: Dehalococcoides sp.
5: Dehalogenimonas lykanthroporepellens
6-7: Roseiflexus sp.
8-9: Roseiflexus castenholzii
10-11: Anaerolinea thermophila
12-14: Chloroflexus sp.
15-17: Chloroflexus aggregans
18-20: Chloroflexus aurantiacus
21-25: Herpetosiphon aurantiacus
10
20

Tab. 1.2. Comparison of mean distances within cyanobacteria and to other eubacterial phyla Comparison of mean distances in the different eubacterial phyla and the 95% confidence intervals of 10,000 mean values calculated from bootstrap samples. Confidence intervals do not overlap between cyanobacteria and the other eubacterial phyla. Distances of 16S rRNA sequences are significantly smaller in cyanobacteria compared to the other prokaryotes. d_W and d_B: mean calculated from the original dataset including all distances. d_W^* and d_B^*: mean of 10,000 means calculated using bootstrap sampling.

Within a genome				
	d_W	d_W^*	**95% confidence intervals**	
			lower	*upper*
Cyanobacteria	0.0001	0.0003	0.0001	0.0005
Chloroflexi	0.0036	0.0020	0.0012	0.0028
Spirochaetes	0.0012	0.0009	0.0005	0.0014
Bacteroidetes	0.0029	0.0023	0.014	0.0032
Between species				
	d_B	d_B^*	**95% confidence intervals**	
			lower	*upper*
Cyanobacteria	0.1427	0.1426	0.1235	0.1587
Chloroflexi	0.3409	0.434	0.2489	0.4087
Spirochaetes	0.3537	0.3541	0.2907	0.4017
Bacteroidetes	0.3779	0.378	0.3390	0.4099

results from the cyanobacteria to three eubacterial phyla (Chroroflexi, Spirochaetes and Bacteroidetes). Figure 1.5 presents the Bayesian consensus phylogenetic tree and the distance matrix reconstructed for the phylum Chloroflexi. Trees and distance matrices for the phyla Spirochaetes, and Bacteroidetes are shown in Additional Files 1.6, 11.7 and 1.8. Within the phylum Chloroflexi, species contain one to five 16S rRNA genes per genome. The phylogenetic tree is well supported by posterior probabilities. Previous phylogenetic studies have divided the phylum Chlorophlexi into several subdivisions [163, 164], the majority of which is supported by our inferred tree. Distances of the 16S rRNA sequences within genomes and between species of Chloroflexi were significantly higher than found for cyanobacteria (Table 1.2). Mean distances of species belonging to the phylum Chloroflexi were $d_W = 0.004$ within species, and showed a 10-fold difference compared to distances between species ($d_B = 0.34$). *Chloroflexus auranticus* and *Chloroflexus sp.* were the only species among the taxa analyzed in this study where 16S rRNA orthologs were more similar than their paralogs. Further comparison of mean distances for 16S rRNA sequences including phyla Spirochaetes and Bacteroidetes confirmed the significantly lower sequence variation in cyanobacteria. A comparison of the distributions of mean distances calculated from the bootstrap re-sampling show no overlap of the 95% confidence intervals of cyanobacteria and any of the other phyla (Additional Files 1.4 and 1.5). Furthermore, within all studied phyla, mean distances for 16S rRNA gene copies within a genome (d_W) were smaller by at least one order of magnitude compared to mean distances for 16S rRNA sequences between species (d_B). Our results support 16S rRNA as an adequate taxonomic marker for the species analyzed in this study and agree with previous findings of limited heterogeneity in 16S rRNA [132].

Evolution of 16S rRNA gene copies in cyanobacteria

Two mechanisms may conserve sequences of gene copies: Purifying selection and concerted evolution. These two can be distinguished by examining variation patterns in non-coding regions [165, 124]. In the case of purifying selection, non-coding regions are thought to evolve neutrally, accumulating mutations over time due to genetic drift. If concerted evolution shapes gene copies, the entire gene sequence including non-coding regions and synonymous sites are homogenized. During this process, genes evolve in 'concert', which is commonly observed in plants and fungi [166, 167] (Figure 1.5). Subsequently, paralogs show stronger similarities than orthologs, as a result of intragenomic homologous recombination [168].

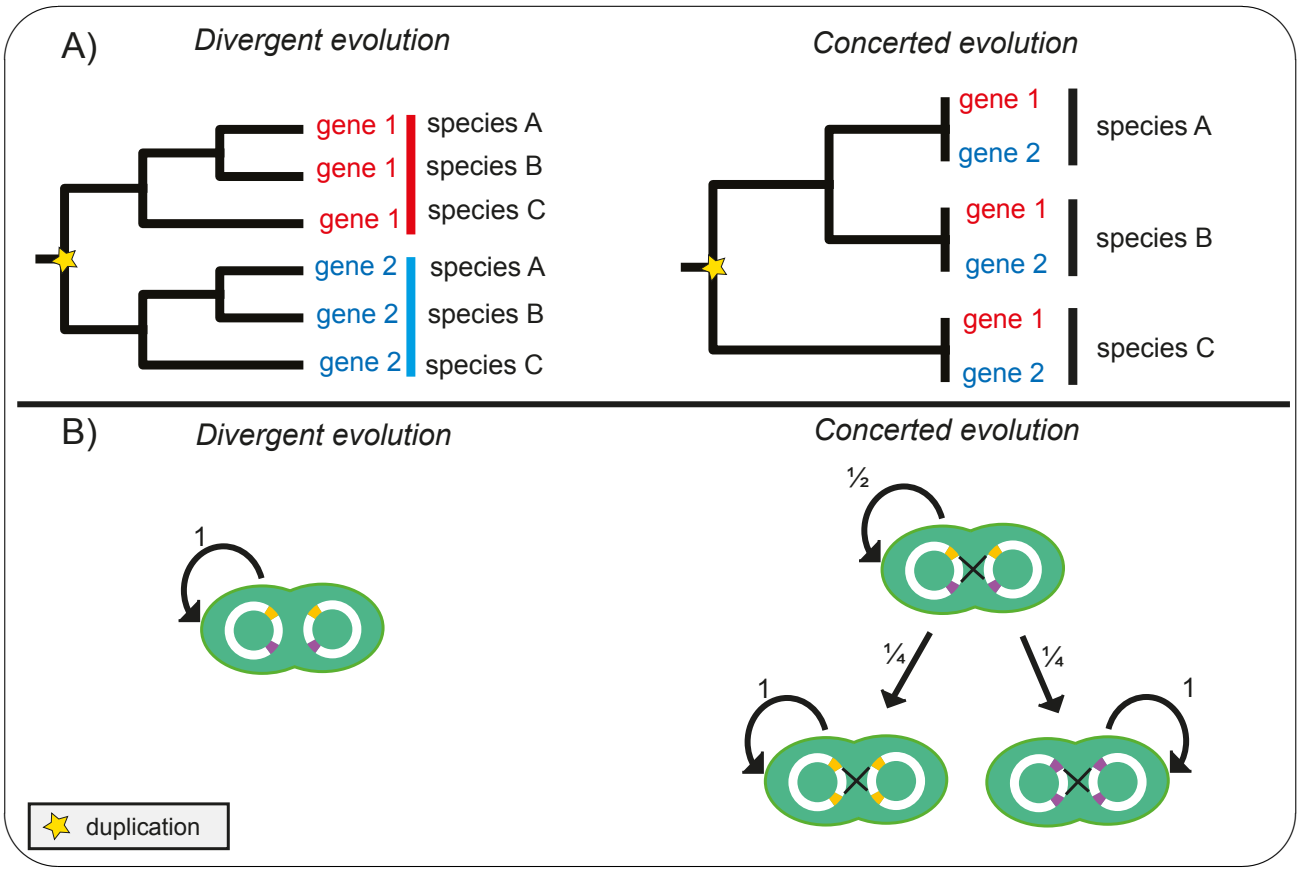


Fig. 1.6. Divergent and concerted evolution. (**A**) The phylogenetic pattern of divergent and concerted evolution evolution. Paralogs and orthologs diverge at similar degrees in the first scenario, while they get frequently homogenized during concerted evolution. A cyanobacterial cell during cell division without homologous recombination. All daughter cells will exhibit the same chromosome as the mother cell. (**B**) Replication pattern during cell division under divergent and concerted evolution. If during cell devision homologous recombination takes place in half of the recombinants the daughter cells will exhibit the same chromosome as the mother. For the other half of recombinants, each gene copy has a $\frac{1}{4}$ chance of replacing the other. Once gene copies are identical homologous recombination cannot reverse the process. Hence if this process is repeated recursively at a population level, one gene copy will eventually get fixed.

The strong conservation of 16S rRNA sequence copies in cyanobacteria and Eubacteria examined here suggests that 16S rRNA in these species is shaped by strong purifying selection or concerted evolution. Generally, it is assumed that ribosomal genes in Archaea and Eubacteria are shaped by concerted evolution [136] (Figure 1.6). 16S rRNA genes can be subdivided in strongly conserved and more variable regions. One would expect

that if purifying selection acts as the major force for conservation of gene copies within a genome, some neutral variation should be detected in these variable regions. The extraordinary conservation of 16S rRNA in cyanobacteria tends to indicate that concerted evolution is a more likely explanation. To verify this suggestion we examined variation in the internal transcribed spacer region, located between the 16S and 23S rRNA gene. The strong conservation found in cyanobacterial 16S rRNA gene sequences could not be confirmed for the ITS-regions of four cyanobacterial taxa (Additional File 1.9). For cyanobacteria and the eubacterial phyla studied here, both concerted evolution and strong purifying selection, appear to be the main contributing factors.

Although, cyanobacteria are assumed to be an ancient phylum which presumably raised oxygen levels in the atmosphere more than 2.3 billion years ago [110], variation in 16S rRNA copies is extremely low. Indeed, phylogenetic tree reconstructions for 16S rRNA result in relatively short estimated branch lengths within this phylum, compared to other eubacterial phyla (Figure 1.2). Short evolutionary distances for 16S rRNA sequences are consistent with a pattern that has been found for morphological characters in cyanobacteria before. In 1994, J.W. Schopf compared the tempo and mode of evolution in cyanobacteria from the Precambrian, to evolutionary patterns observed in fossils during the Phanerozoic. The latter have been described by G.G. Simpson in his book “The tempo and mode of evolution" [169]. Schopf found that evolutionary predictions which Simpson made for metazoan fossils from the Phanerozoic, can also be applied to cyanobacteria. Morphologically, cyanobacteria seem to evolve not only at a “bradytelic", but “hypobradytelic” mode, meaning at exceedingly low evolutionary rates. Fossils from the Precambrian strongly resemble present morphotypes. The oldest undisputed cyanobacterial fossils date back circa 2.0 billion years [103, 123]. Morphological appearance of these microfossils already suggests the presence of at least four of the morphological sections described by Castenholz [64]. It seems that cyanobacteria reached their maximum morphological complexity two billion years ago, and many of today’s species could be described as so-called ‘living fossils’. Although, these observations have not been confirmed for biochemical or genomic features, yet, it appears from our analyses that 16S rRNA genes fulfill criteria of exceptionally slow evolutionary rates when compared to other prokaryotes.

Conclusion

Among 22 fully sequenced cyanobacterial taxa that were carefully chosen according to phylogenetic position and morphological characteristics, we identified 41 protein coding genes that occur as multiple highly conserved copies in at least one cyanobacterial species. Copy numbers of ribosomal genes show a significant correlation to cyanobacterial species that are capable of terminal differentiation. The formation of heterocysts, morphologically modified cells for nitrogen fixation, requires a strong increase in gene expression, for which an accumulation of ribosomes could be of potential advantage. Further testing would be required though, to make causal conclusions for increased rRNA operons in cyanobacteria belonging to section IV and V. Furthermore, phylogenetic analyses revealed a high conservation of 16S rRNA copies within eubacterial species. Though this is true for all phyla that have been analyzed, cyanobacteria exhibit an exceptionally strong con-

servation. Comparison to variation in ITS regions point to concerted evolution via homologous recombination and purifying selection as forces behind 16S rRNA sequence evolution. Comparison of interspecific genetic distances within several prokaryotic phyla, showed significantly lower variation of cyanobacterial 16S rRNA gene sequences. This suggests that 16S rRNA gene sequences evolve by a 'hypobradytelic' mode of evolution, previously suggested for morphological characteristics in cyanobacteria [106].

Methods

Data choice and description

For this study we only used cyanobacterial taxa with fully sequenced and annotated genomes publicly available on GenBank ($http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi$). Of those 42 genomes (as of August 2011), 36 belong to singlecelled strains, covering 10 different species in total. The remaining six genomes belong to multicellular strains, each representing another species. The taxon sampling was done to exclude a bias towards unicellular closely related cyanobacteria which are overrepresented in the genome-database [170]. Therefore, to cover the widest possible range of morphotypes, we selected one or more, fully sequenced taxa of each species for a total dataset of 22 cyanobacterial strains. More precisely, we included multiple strains of species *Cyanothece sp.*(2), *Synechococcus sp.*(4), and *Prochlorococcus marinus*(3), which, following the examination of previous phylogenies [171, 172, 120, 159], are assumed to add phylogenetic diversity. No outgroup was included in the phylogenetic analyses. *Gloeobacter violceus* has been shown to be closest to eubacterial outgroups [159]. Therefore, phylogenetic trees are represented accordingly.

Identification of conserved paralogs and correlation to morphotypes

In order to find genes with multiple copies, we applied the orthology prediction algorithm OMA [173] to the set of 22 complete cyanobacteria genomes. First we looked for clusters of highly conserved paralogous genes within each species. From the all-against-all pairwise sequence alignments computed by OMA, we selected pairwise hits within each species with an alignment score of at least 130 and minimum sequence identity of $\geq$ 98%, $\geq$ 95% and $\geq$ 90%. We then used these hits as edges in a homology graph, and identified clusters of highly conserved paralogs as connected components. Finally, we removed hits within a cluster if the pairwise distance differed significantly from the mean distance within the cluster. In the second step, we grouped detected homologous clusters across species using OMA alignments, but this time with a score cut-off of 180 and minimum sequence identity of $\geq$ 50%. We further required that $\geq 0.8 \cdot n_i \cdot n_j$ of hits between any pair of clusters i and j be present in order to be considered, where n_i, n_j is the number of genes in clusters i and j, respectively. If a cluster in one genome grouped with several clusters in another genome, we chose the one with the lowest average pairwise distance. Again, homologous groups were extracted as connected components from the resulting graph. Finally, single orthologs from the OMA orthologous matrix (i.e, with no

detected multiple copies within their originating genome) were matched and added to corresponding homologous groups.

We tested whether a correlation between cell differentiation and copy numbers could be observed for the identified genes. To do this, we devided cyanobacterial species into four different groups of cell differentiation (G0-G3; see results). Five strains belong to G0, 12 taxa belong to G1, *Tricodesmium* is the only genus in G2, and four species belong to G3. For 16S rRNA genes additional data could be obtained from rrndb-database [161] (Additional File 1.3). Adding these data resulted in a taxon set of 16S rRNA gene sequences as follows: five strains belonging to G0, 12 strains representing G1, *Trichodesmium* as the only species in G2 and 11 species in G3. Spearman's rank and Pearson's correlation coefficients were applied in order to estimate associations between conserved copy numbers and morphological groups (G0-G3), using R-software. Correlations with a p-value$<$ 0.01 were considered to be significant.

Phylogenetic analyses

We conducted separate phylogenetic analyses of 16S rRNA gene sequences of cyanobacteria (Table 1.1) and four different eubacterial phyla (Additional File 1.10). For all taxa included in the phylogenetic trees, full genome sequences were available. All sequences were downloaded from GenBank [174]. For cyanobacteria two phylogenetic trees were reconstructed. One including a single 16S rRNA sequence per taxon and another including all 16S rRNA copies per taxon. Final taxon sets included 22 sequences in the first case and 48 sequences in the latter. The datasets were aligned using Clustal-X software with default settings [175] (1,325nt incl. gaps). Gaps were excluded from the analysis. Phylogenetic reconstructions were done using Bayesian analysis as implemented in MrBayes software [176]. Two Metropolis coupled Markov Chain Monte Carlo (MC^3) searches were run for 10^7 generations each using three heated and one cold chain. Figures 1.1 and 1.2 show the consensus trees of 16,002 trees that were sampled every 1,000th generation from the MC^3 searches, excluding the first 2,000 trees of each run (burn-in). At that point the log probabilities reached stationarity and average standard deviation of split frequencies were below 0.02. Performance of the MCMC and stationarity of the parameters were checked using Tracer v1.5 [177]. Effective Sample Sizes (ESS) were all above 200, supporting a well mixed MCMC run.

Phylogenetic analysis described for cyanobacteria was equally conducted for the phyla Auificae, Bacteroidetes, Chloroflexi and Spirochaetes. The non-cyanobacterial phylogenetic trees were reconstructed including all 16S rRNA gene copies of each taxon. MC^3 analyses were run for 10^6 generations. The first 200,000 generations of each run were discarded as a burn-in. Parameters and trees were sampled every 1,000th generation resulting in a final set of 1,602 trees. The resulting Bayesian consensus trees for each phylum with posterior probabilities displayed at the nodes, have been visualized with FigTree v1.3.1 [178].

Molecular distance analyses

For each set of aligned 16S rRNA gene sequences, distance matrices were calculated applying a K80 substitution model as implemented in the program baseml of PAML v4.3 [179]. The same was done for the internal transcribed spacer region (ITS) in cyanobacteria (Additional File 1.9). The resulting numeric matrices were imaged as color matrices using the R-package “plotrix" [180]. The color gradient of each matrix was scaled by the matrix's minimum and maximum values. Mean distances were calculated within strains (between paralogs; d_W) and between strains (between orthologs; d_B), for each phylum. Significant differences in mean distances were confirmed with bootstrap re-samplings of independent values from the original dataset. To estimate significant differences of mean distances within species (d_W), independent distance values were sampled 10,000 times for each species. Bootstrap re-sampling was done on each of these sample sets. Mean distances were hence calculated and their distribution plotted in a histogram (Additional File 1.4). The resulting overall mean, d_W^* of the distributions, as well as 95% confidence intervals are presented in Table 1.2. To confirm potential differences of mean distances between species (d_B) compared to other phyla, independent values were sampled 10,000 times. These datasets were re-sampled and mean distances calculated. The distributions are displayed in Additional File 1.5. The resultant overall mean, d_B^* of each distribution, as well as 95% confidence intervals are shown in Table 1.2. Independence of distance estimations was assumed if from the corresponding matrix each column and row was only chosen once.

Authors contributions

BES and HCB conceived the study; BES gathered data; BES and DAD conducted analyses; BES, DAD, MA and HCB designed research and wrote the paper. All authors read and approved the final manuscript.

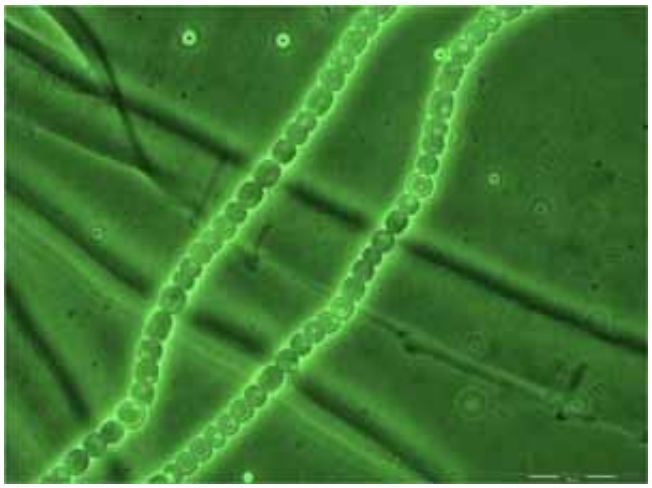
Acknowledgements

For statistical advice and support we would like to thank Erik Postma. Furthermore, we would like to thank Dr. Manuela Filippini Cattani, Dr. Miroslav Svercel and Valentina Rossetti for helpful comments on various versions of the manuscript.

Competing interest

The authors declare that they have no competing interests.

CHAPTER II



The origin of multicellularity in cyanobacteria

Authors:

Bettina E. SCHIRRMEISTER, Alexandre ANTONELLI and Homayoun C. BAGHERI

Schirrmeister et al. (2011) BMC Evol. Biol. 11:45. -ranked #5 of most viewed BMC Evol.Biol. articles in the past year

Background: Cyanobacteria are one of the oldest and morphologically most diverse prokaryotic phyla on our planet. The early development of an oxygen-containing atmosphere approximately 2.45 - 2.22 billion years ago is attributed to the photosynthetic activity of cyanobacteria. Furthermore, they are one of the few prokaryotic phyla where multicellularity has evolved. Understanding when and how multicellularity evolved in these ancient organisms would provide fundamental information on the early history of life and further our knowledge of complex life forms.

Results: We conducted and compared phylogenetic analyses of 16S rDNA sequences from a large sample of taxa representing the morphological and genetic diversity of cyanobacteria. We reconstructed ancestral character states on 10,000 phylogenetic trees. The results suggest that the majority of extant cyanobacteria descend from multicellular ancestors. Reversals to unicellularity occurred at least 5 times. Multicellularity was established again at least once within a single-celled clade. Comparison to the fossil record supports an early origin of multicellularity, possibly as early as the "Great Oxygenation Event" that occurred 2.45 - 2.22 billion years ago.

Conclusions: The results indicate that a multicellular morphotype evolved early in the cyanobacterial lineage and was regained at least once after a previous loss. Most of the morphological diversity exhibited in cyanobacteria today —including the majority of single-celled species— arose from ancient multicellular lineages. Multicellularity could have conferred a considerable advantage for exploring new niches and hence facilitated the diversification of new lineages.

Background

Cyanobacteria are oxygenic phototrophic prokaryotes from which chloroplasts, the light harvesting organelles in plants, evolved. Some are able to convert atmospheric nitrogen into a form usable for plants and animals. During earth history, cyanobacteria have raised atmospheric oxygen levels starting approximately 2.45 - 2.22 billion years ago and provided the basis for the evolution of aerobic respiration [181, 110, 182, 143, 183, 184, 111]. Cyanobacteria have also evolved extensive morphological diversity. Various patterns of cell organization exist, ranging from single-celled to differentiated multicellular forms with branching patterns. Species of this phylum occupy various habitats. They can be found in marine, freshwater or terrestrial environments, ranging from polar to tropical climate zones. Based on their morphology, they have been divided into five sections [84, 64] (Table 2.1). Sections I and II comprise single-celled bacteria, whereas sections III to V comprise multicellular forms. The latter sections are distinguished according to their level of organization. Section III is multicellular and undifferentiated, sections IV and V are multicellular and differentiated. The latter have the ability to produce heterocysts for nitrogen fixation and akinetes (climate-resistant resting cells). In addition, species in section V have the ability to branch in multiple dimensions.

Different interpretations of multicellularity are currently used [185, 186, 187]. For cyanobacteria, characterization of multicellularity has been described in previous studies [188, 189, 190, 191]. Cell to cell adhesion, intercellular communication, and for more complex species, terminal cell differentiation seem to be three essential processes that define multicellular, prokaryotic organisms on this planet [191]. Some forms of complexity found in several multicellular eukaryotes are not present in prokaryotes, but simple forms of multicellularity can be identified in three sections of the phylum cyanobacteria. Multicellular patterns comprise basic filamentous forms as found for section III, as well as more complex forms involving terminal differentiation, present in sections IV and V. In eukaryotes, multicellular complexity ranges from what is comparable to cyanobacteria to cases with up to 55 cell types as estimated for higher invertebrates such as arthropods or molluscs [192]. Considering that cyanobacterial sections III, IV and V resemble some of the first forms of multicellular filaments on Earth, knowing when and how these shapes evolved would further our understanding of complex life forms.

Some of the oldest body fossils unambiguously identified as cyanobacteria have been found in the Kasegalik and McLeary Formations of the Belcher Subgroup, Canada, and are evaluated to be between 1.8 billion and 2.5 billion years old [103, 184]. Studies from $\sim$2.0 billion year old formations [103, 123] contain both unicellular and multicellular morphotypes of cyanobacteria. Cyanobacteria certainly existed as early as 2.32 billion years ago, if one accepts the assumption that they were responsible for the rapid accumulation of oxygen levels, known as the "Great Oxygenation Event" [181, 110, 182, 183, 111]. Multicellular fossils belonging to the cyanobacteria are well known from the late Precambrian [106, 187, 193] and possibly already existed 2.32 billion years ago. Other microbe-like multicellular filaments even older than 3.0 billion years have been found several times [194, 195, 108, 196, 197]. Some of the latter fossils are morphologically similar to species from the cyanobacterial order Oscillatoriales [105, 198], but no clear evidence has been adduced yet. Although

Tab. 2.1. Subset of cyanobacterial taxa used for the analyses with GenBank accession numbers for 16S rDNA sequences.

unicellular strains	**accession numbers**	**multicellular strains**	**accession numbers**
Section I		**Section III**	
Chamaesiphon subglobosus PCC 7430[1]	AY170472	*Arthronema gygaxiana* UTCC 393	AF218370
Cyanobium sp. JJ23-1	AM710371	*Arthrospira platensis* PCC 8005	X70769
Cyanothece sp. PCC 8801[1]	AF296873	*Crinalium magnum* SAG 34.87	AB115965
Chroococcus sp. JJCM	AM710384	*Filamentous thermophilic cyanobacterium*	DQ471441
Dactylococcopsis sp.[1]	AJ000711	*Geitlerinema sp.* BBD HS217[1]	EF110974
Gloeobacter violaceus PCC 7421[1]	BA000045	*Halospirulina sp.*[1]	NR_026510
Gloeothece sp. PCC 6909/1[1]	EU499305	*Leptolyngbya sp.* ANT.LH52.1	AY493584
Microcystis aeruginosa strain 038[1]	DQ363254	*Lyngbya aestuarii* PCC 7419[1]	AB075989
Prochlorococcus sp. MIT9313[1]	AF053399	*Microcoleus chthonoplastes* PCC 7420[1]	AM709630
Prochloron sp.[1]	X63141	*Oscillatoria sp.*[1]	AJ133106
Radiocystis sp. JJ30-3	AM710389	*Oscillatoria sancta* PCC 7515	AF132933
Synechococcus elongatus PCC 6301[1]	AP008231	*Phormidium mucicola* IAM M-221	AB003165
Synechococcus sp. CC9605	AY172802	*Plectonema sp.* F3[1]	AF091110
Synechococcus sp. WH8101	AF001480	*Planktothrix sp.* FP1	EU078515
Synechocystis sp. PCC 6803	NC_000911	*Prochlorothrix hollandica*[1]	AJ007907
Synechocystis sp. PCC 6308[1]	AB039001	*Pseudanabaena sp.* PCC 6802	AB039016
Synechocystis sp. CR_L29[1]	EF545641	*Pseudanabaena sp.* PCC 7304 [1]	AF132933
Synechococcus sp. P1	AF132774	*Spirulina sp.* PCC 6313	X75045
Synechococcus sp. C9[1]	AF132773	*Starria zimbabweensis* SAG 74.90[1]	AB115962
Synechococcus lividus C1	AF132772	*Symploca sp.*PCC 8002	AB039021
Acaryochloris sp. JJ8A6[1]	AM710387	*Trichodesmium erythraeum* IMS 101[1]	AF013030
Thermosynechococcus elongatus BP-1[1]	BA000039	**Section IV**	
Section II		*Anabaena sp.* PCC 7108	AJ133162
Chroococcidiopsis sp. CC2	DQ914864	*Calothrix sp.* PCC 7103[1]	AM230700
Dermocarpa sp. MBIC10768	AB058287	*Nodularia sp.* PCC 7804[1]	AJ133181
Dermocarpella incrassata	AJ344559	*Nostoc sp.* PCC 7120	X59559
Myxosarcina sp. PCC 7312[1]	AJ344561	*Scytonema sp.* U-3-3[1]	AY069954
Myxosarcina sp. PCC 7325	AJ344562	**Section V**	
Pleurocapsa sp. CALU 1126	DQ293994	*Chlorogloeopsis sp.* PCC 7518[1]	X68780
Pleurocapsa sp. PCC 7516	X78681	*Fischerella sp.* PCC 7414	AB075986
		Symphyonema sp. strain 1517	AJ544084
		Eubacteria	
		Beggiatoa sp. 'Chiprana'	EF428583

biogenicity of some of the oldest fossils has been questioned [199, 200], a large variety of bacteria including anoxic phototrophs already existed by the time cyanobacteria evolved oxygenic photosynthesis[197]. Though impressive for prokaryotes, the fragmentary fossil record alone is not sufficient to disentangle the origin of cyanobacteria and their morphological phenotypes. Therefore, additional methods such as phylogenetic analysis provide a promising possibility to gather further clues on the evolution of such a complex phylum.

Phylogenetic analyses of cyanobacteria have gained in quantity over the past 20 years [112, 115, 85, 116, 201, 141, 118, 143, 119, 120]. These studies have shown that morphological characterization does not necessarily reflect true relationships between taxa, and possibly none of the five traditional morphological sections is monophyletic. Similar morphologies must have evolved several times independently, but details on this morphological evolution are scarce. Analyses assessing characteristics of cyanobacterial ancestors [118, 120]

provide not only fundamental information on the history of cyanobacteria, but also on the evolution of life forms in the Archean Eon.

If one studies phylogenetic relationships based on protein coding genes in bacteria, it is possible to encounter the outcome of horizontal gene transfer (HGT) [202]. This issue is not as problematic for ribosomal DNA [203]. Nonetheless, the problem could be potentially reduced by analyzing datasets of concatenated conserved genes. Identification of these genes for phylogenetic analyses is not without difficulty, and requires in an ideal case comparison of complete genome data [204]. In cyanobacteria, many phylogenetic studies have concentrated on specific clades or smaller subsets of known species in this diverse phylum [205, 206, 171, 207, 208, 209, 120]. Therefore the genomic data presently available are strongly biased towards certain groups. In particular, genomic studies in cyanobacteria have emphasized marine species from Section I. Marine microphytoplankton (*Synechococcus* and *Prochlorococcus*) are a particularly well studied group [205, 171, 208, 209], reflected by 19 sequenced genomes out of 41 cyanobacterial genomes sequenced to date (*http : //www.ncbi.nlm.nih.gov/genomes/lproks.cgi*, accessed in January 2011). From species belonging to section III only two genomes (*Trichodesmium erythraeum* and *Arthrospira platensis) are* known. For sections IV (four genomes known) and V (no genomes known) molecular data are rare or missing. As genomic data accumulate, promising phylogenomic approaches to cyanobacteria are being established [118, 208, 119, 120]. Despite these advances, it is at present difficult to obtain sequences other then 16S rDNA to cover a representative sample of species from all five sections.

The aim of this paper is to use molecular phylogenetic methods to address the evolutionary history of cyanobacteria and the evolution of multicellularity. For this purpose, we established a phylogeny based on 16S rDNA sequences belonging to 1,254 cyanobacterial taxa. From that phylogeny we sampled 58 cyanobacterial taxa that represent all main clades obtained and all five sections described by Castenholz *et al.* [84, 64], and feature a 1:1 ratio of unicellular to multicellular species. We used several methods to reconstruct the morphological evolution of ancestral lineages, and compared our results to known fossil data. Since the fossil record is inconclusive on the timing and taxonomic position of multicellular cyanobacteria, our study provides independent evidence on the first appearance and evolution of multicellularity among the ancestors of living cyanobacteria.

Results and Discussion

Phylogenetic analysis

Phylogenetic analyses of all identified cyanobacteria

To infer the evolution of multicellularity in cyanobacteria we carried out several phylogenetic analyses. To ensure a correct taxon-sampling, a phylogeny containing 1,254 16S rDNA sequences of cyanobacteria obtained from GenBank was reconstructed (Figure 2.1). Cyanobacterial morphotypes were assigned to four groups (A-D) which correlate to the five sections described by Castenholz *et al.* [64]. Using this nomenclature, sub-

groups in the phylogeny were assigned to one of the four different morphological groups (A-D) according to their dominant shape. In total 14 sub-groups were identified for the phylogenetic tree. Five sub-groups consist of unicellular species from section I (A1-A5), two sub-groups are composed of single celled section II bacteria (B1,B2), four sub-groups are made up of multicellular species belonging to section III (C1-C4) and two sub-groups cover differentiated species from section IV and V (D1-D2). One sub-group contains both species from section I and III and is therefore designated as AC1. The phylogeny further contains six chloroplast genomes from the eukaryotic phyla Glaucophyta, Rhodophyta and the division Chlorophyta. Chloroplast sequences branch close to the bases and form a sister group to the cyanobacterial sub-groups mentioned. Furthermore six different Eubacteria were included in the phylogeny. They appear to form a distinct outgroup to the cyanobacteria and chloroplasts.

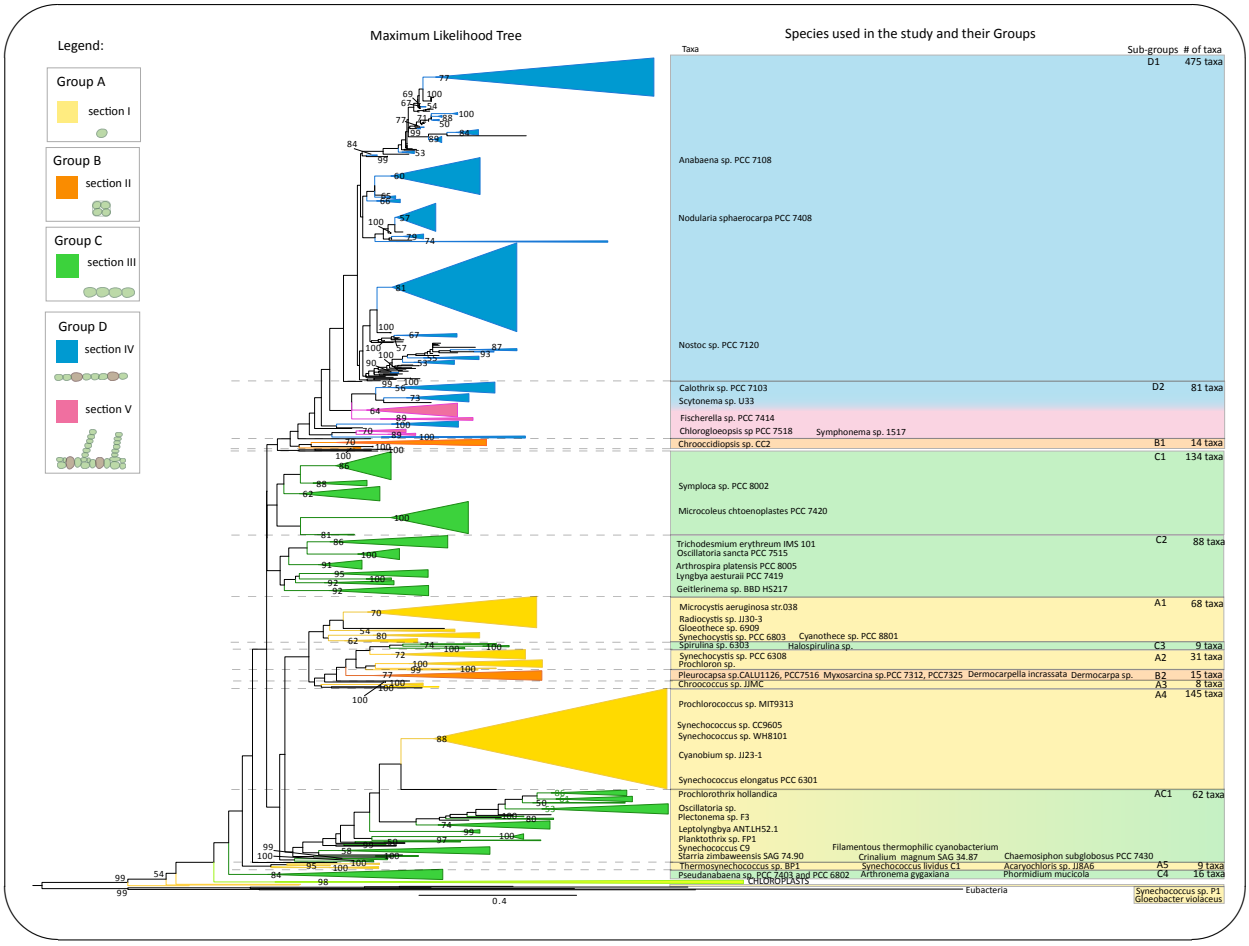


Fig. 2.1. Phylogenetic tree of 1,254 cyanobacterial species. Maximum likelihood phylogram of cyanobacteria, based on GTR+G+I substitution model. Six eubacterial species form an outgroup. The ingroup contains 1,254 cyanobacterial strains and six different chloroplast sequences. Bootstrap values (> 50%) calculated from 100 re-samplings are displayed at the nodes. Colors define major morphological characters in the groups. Yellow are single-celled cyanobacteria of section I; orange single-celled from section II; green are multicellular, undifferentiated cyanobacteria from section III; blue are multicellular and differentiated bacteria from section IV; and pink from section V. Sections as described by Castenholz 2001 [64]. Different sub-groups (AC1;A1-A5;B1,B2;C1-C4;D1-D2) are defined for the phylogeny. Nomenclature of groups correlates with morphological sections as illustrated in the legend. From these sub-groups taxa were sampled for further analyses. A complete list with species included in the analysis can be found in Additional File 2.1. The newick format of the tree is given in Additional File 2.2.

Phylogenetic analyses to identify an outgroup

Rooted and unrooted phylogenetic analyses reconstructed with maximum likelihood and Bayesian inference and based on 16S rRNA gene sequences of 27 eubacterial species (Table 2.2), including 5 cyanobacteria revealed congruent results. Cyanobacteria form a monophyletic group. Figure 2.2 shows the unrooted Bayesian consensus tree which supports cyanobacterial monophyly with posterior probabilities (PP)/bootstrap values (BV) of 1.0/100%. Phylogenetic trees constructed with an archaean outgroup support cyanobacterial monophyly with PP/BV of 1.0/98% (Additional File 2.3). In both cases, *Plantomyces brasilienses* and *Chlamydia trachomatis*, both gram negative bacteria, form a sister group to the cyanobacteria. This does not agree with other studies [210, 211, 212, 213], where *Deinococcus-thermus* was suggested to be the closest eubacterial relative to cyanobacteria. These discrepancies may be due to a lack of information when solely using 16S rRNA gene sequences for such distant relations. Furthermore, our results confirm the basal position of *Gloeobacter violaceus*, closest to the rest of the eubacteria, as found elsewhere [212]. This supports previous findings which state that *Gloeobacter violaceus* diverged very early from cyanobacteria living today [214, 115, 85, 215]. *Gloeobacter* shows differences in cell structure and metabolism that clearly distinguish it from the rest of extant cyanobacteria [216, 160]. It lacks thylacoid membranes and many genes from Photosystems I and II. Phylogenetic relations of the other eubacterial species show only weak support and are therefore not discussed further.

Tab. 2.2. Non-cyanobacterial species used in this study with GenBank accession numbers for 16S rDNA sequences. [1] *Taxonomy as described at http://www.ncbi.nlm.nih.gov/Taxonomy/ and [212].*

	Phyla[1]	**species**	**accession numbers**
EUBACTERIA	Actinobacteria	*Actinosynnema mirum* DSM 43827	CP001630
	Aquificae	*Aquifex aeolicus* VF5	NC_000918
	Bacteroidetes/Chlorobi	*Bacteroidetes bacterium* X3-d	HM212417
	Bacteroidetes/Chlorobi	*Chlorobium sp.* sy9	EU770420
	Chlamydiae/Verrucomicrobia	*Chlamydia trachomatis*	AM884176
	Chlamydiae/Verrucomicrobia	*Verrucomicrobia bacterium* YC6886	FJ032193
	Chloroflexi	*Chloroflexus sp.* Y-400-fl	NC_012032
	Chrysiogenetes	*Chrysiogenes arsenatis*	NR_029283
	Deferribacteres	*Deferribacter desulfuricans* SSM1	AP011529
	Deinococcus-Thermus	*Deinococcus sp.* AA63	AJ585986
	Dictyoglomi	*Dictyoglomus turgidum*	NC_011661
	Fibrobacter/Acidobacteria group	*Acidobacterium capsulatum* ATCC 51196	CP001472
	Fibrobacter/Acidobacteria	*Fibrobacter succinogenes*	NC_013410
	Firmicutes	*Streptococcus mutans* NN2025	AP010655
	Fusobacteria	*Fusobacterium nucleatum*	GU561358
	Gemmatimonadetes	*Gemmatimonas sp.*	GU557153
	Nitrospira	*Nitrospira calida*	HM485589
	Planctomycetes	*Planctomyces brasiliensis* DSM 5305	CP002546
	Proteobacteria	*Beggiatoa sp.* 'Chiprana'	EF428583
	Spirochaetes	*Spirochaeta thermophila* DSM 6192	NC_014484
	Thermodesulfobacteria	*Thermodesulfobacterium hydrogeniphilum*	AF332514
	Thermotogae	*Thermotoga lettingae* TMO	NC_009828
ARCHAEA	Nanoarchaeota	*Nanoarchaeum equitans* Kin4-M chromosome,	NC_005213

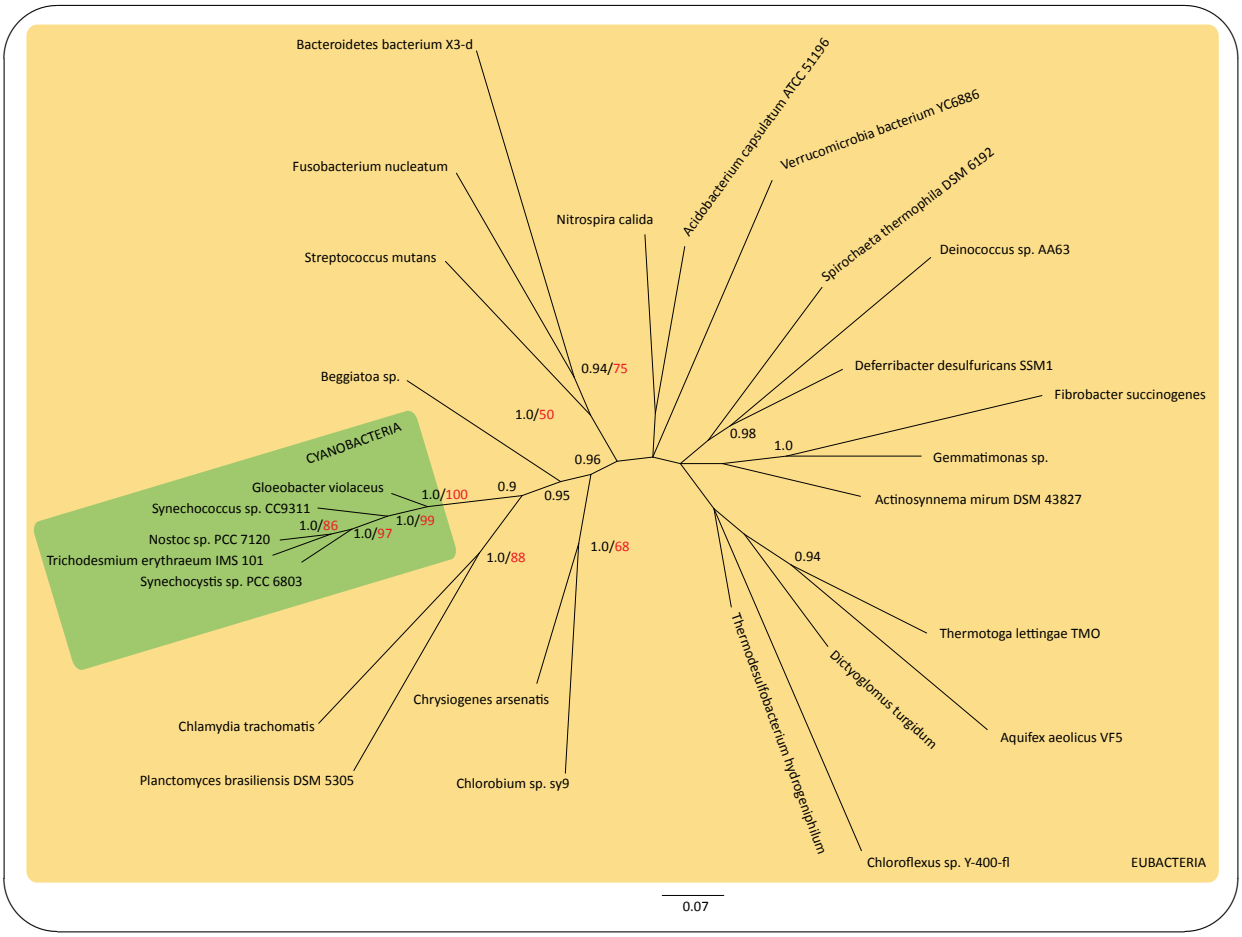


Fig. 2.2. Unrooted Bayesian consensus tree of eubacteria including five cyanobacterial species. phylogenetic tree of 16S rRNA gene sequences from 27 eubacterial species reconstructed using Bayesian methods. Posterior probabilities (black) and bootstrap values (red) from 100 resamplings are displayed at the nodes. Cyanobacteria, represented by 5 species, form a monophyletic group with *Gloeobacter violaceus* being closest to other eubacterial species.

We separately tested each of the 22 eubacterial species (Table 2.2) originating from a diverse set of non-cyanobacterial phyla, with a subset of the cyanobacteria (58 taxa). The latter were chosen from the large dataset containing 1,254 taxa, and cover all sub-groups of the tree. This subset was used for all subsequent phylogenetic analyses. Though multicellular species seem to dominate the known cyanobacteria, we chose to sample a taxa set containing unicellular and multicellular morphotypes in a 1:1 ratio. That way biases towards certain character states would be excluded. Furthermore, taxa used in the analyses should represent species from all five sections described by Castenholz *et al.* [64]. Given our interest in the base of the phylogeny, a greater number of taxa were sampled from basal sub-groups. Due to a lack of data available on GenBank at the present state of research, efforts to build a phylogenetic reconstruction of this size (58 species) using additional ribosomal protein sequences failed. But genomic data are accumulating (57 genomes in progress according to GenBank) and will soon offer possibilities for further extensive analyses.

Results of six phylogenetic trees are displayed in Figure 2.3 (Additional file 2.4: Newick format of all trees). The majority of the trees exhibit a topology that agrees with Figure 2.2, with the position of *Gloeobacter violaceus* close to the outgroup. Strong differences are found in group support within the trees. In 14 of the 22 trees,

three nodes could be identified which lead to three clades, named here E (**E**ntire five sections(A-D)), AC and C (nomenclature as described for the large tree; Figure 2.1). *Gloebacter violaceus* and *Synechococcus P1* are found at the base of the cyanobacterial phylogeny in 16 trees, from which 7 trees exhibit *Gloeobacter violaceus* closest to the eubacterial outgroup.

In total 14 trees showed congruent topologies. From the 14 eubacteria which have been used as an outgroup in these trees, we chose *Beggiatoa sp.* as an outgroup for further analyses because its 16S rRNA gene sequence exhibits the shortest distance to the cyanobacteria.

Phylogenetic analyses of a cyanobacterial subset

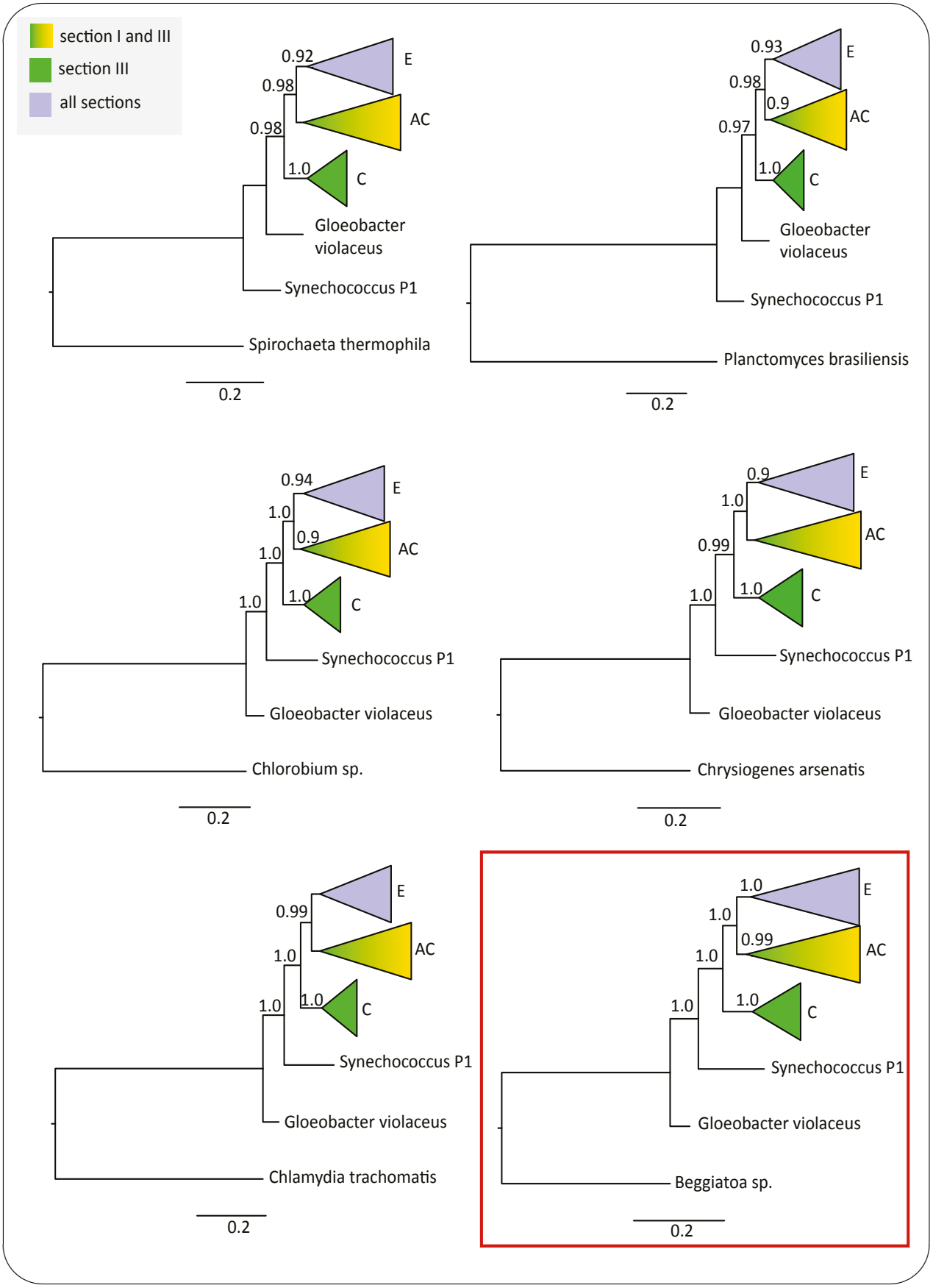
Phylogenetic analyses of 16S rRNA gene sequences from a subset of 58 cyanobacterial taxa were conducted using maximum likelihood (Additional File 2.6) and Bayesian inference (Figure 2.4). For taxa that diverged a long time ago, there is a possibility of sequence saturation, in which case further mutations would have no effect on the distance between sequences any more. We could significantly reject the possibility of sequence saturation for our alignment (Additional File 2.5).

A general substitution model (GTR+G+I) was applied for both analyses. Results of the maximum likelihood and Bayesian methods are highly congruent.

Result of the Bayesian analysis with posterior probabilities (black) and bootstrap values (red) displayed at the nodes is pictured in Figure 2.4. Posterior probabilities above 0.95 and bootstrap values over 70% are considered to represent a high phylogenetic support. Bootstrap values between 50% and 70% are considered weak support. Posterior probabilities below 0.90 and bootstrap values below 50% are not displayed. At deep nodes, the tree topology is fully resolved with high posterior probabilities. Apart from section V, none of the morphological sections described by Castenholz *et al.* [64] is monophyletic. Compared to the outgroup *Beggiatoa sp.*, branch lengths are relatively short, which seems surprising given the old age of the phylum. Rates of evolution in cyanobacteria are extremely slow. This so called "hypobradytelic" tempo would explain their short evolutionary distances [106, 217, 218].

Cyanobacteria form three distinct clades mentioned earlier (Figure 2.2). Clades E , AC and C exhibit posterior probabilities (PP)/bootstrap values (BV) of 1.0/51%, 0.99/-, and 1.0/97% respectively (no support: "-"). Clade E comprises all taxa analyzed from section II, some from section I (*Synechocystis, Microcystis, Gloeothece* and others), some from section III (*Oscillatoria, Trichodesmium, Arthrospira, Lyngbya, Microcoleus, Spirulina* and others) and all from sections IV and V. Within clade E two subclades, E1 (species from section II; PP/BV=1.0/81%) and B (species from sections IV and V among others; PP/BV=1.0/100%), are found. Clade AC contains species from section I and III (among others, species from the genera *Synechococcus, Prochlorococ-*

Fig. 2.3 *(on the next page).* Bayesian consensus trees of cyanobacterial subset using different eubacterial outgroups. Six out of 22 phylogenetic trees reconstructed with Bayesian inference. For each tree an outgroup from a different eubacterial phylum was chosen. Posterior probabilities are displayed at the nodes. Green color represents multicellular cyanobacteria from section III, green-yellow gradient covers species from unicellular section I and multicellular section III, and purple depicts all five different morphological sections present in cyanobacteria. The majority of outgroups exhibits a similar tree topology. For further analyses *Beggiatoa sp.* was selected as an outgroup.


section I and III
section III
all sections
0.92
E
0.98
AC
0.98
1.0
C
Gloeobacter violaceus
Synechococcus P1
Spirochaeta thermophila
0.2
0.93
E
0.98
0.9
AC
0.97
1.0
C
Gloeobacter violaceus
Synechococcus P1
Planctomyces brasiliensis
0.2
0.94
E
1.0
0.9
AC
1.0
1.0
C
1.0
Synechococcus P1
Gloeobacter violaceus
Chlorobium sp.
0.2
0.9
E
1.0
AC
0.99
1.0
1.0
C
Synechococcus P1
Gloeobacter violaceus
Chrysiogenes arsenatis
0.2
0.99
E
AC
1.0
1.0
1.0
C
Synechococcus P1
Gloeobacter violaceus
Chlamydia trachomatis
0.2
1.0
E
1.0
0.99
AC
1.0
1.0
1.0
C
Synechococcus P1
Gloeobacter violaceus
Beggiatoa sp.
0.2

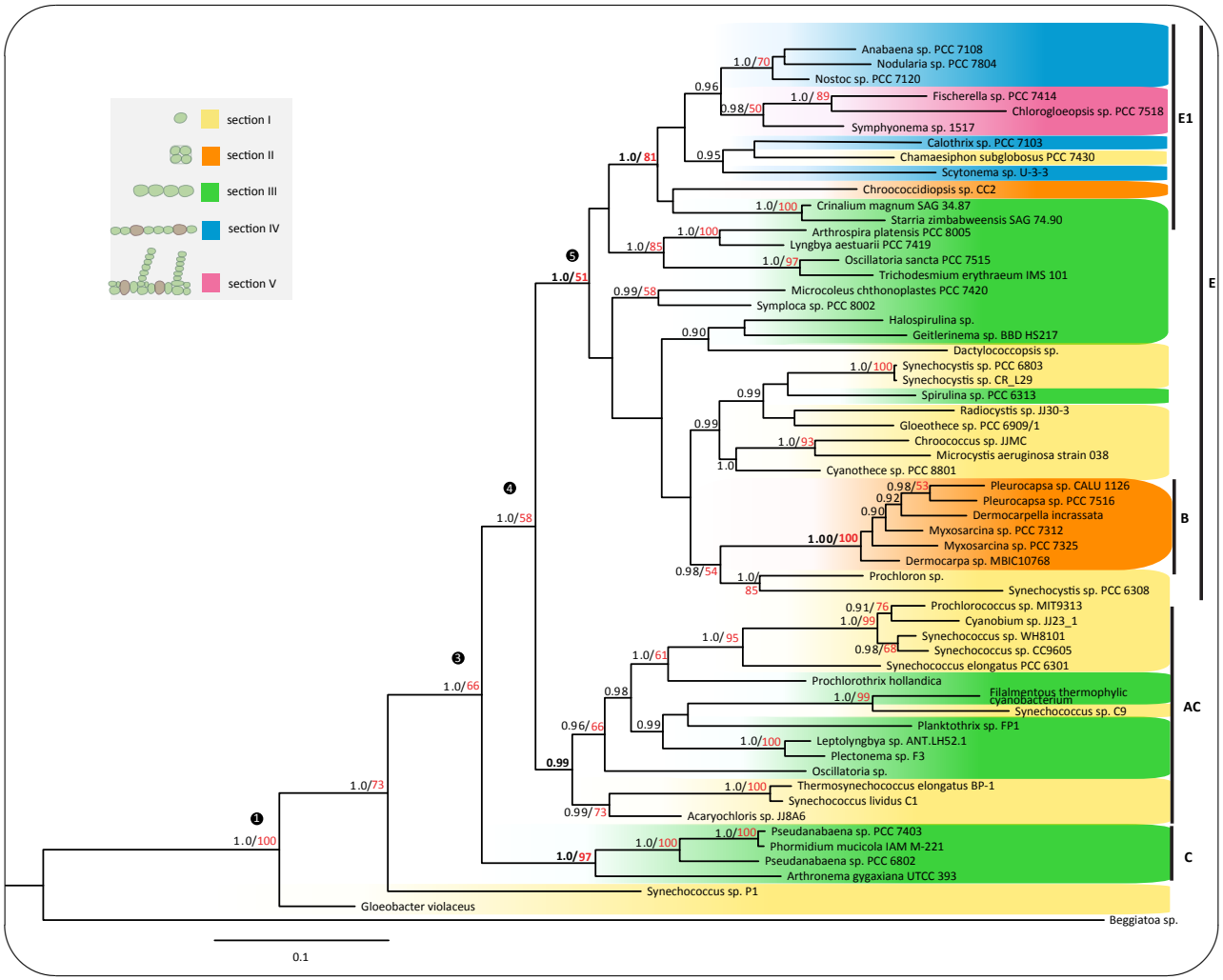


Fig. 2.4. Phylogenetic tree of a cyanobacterial subset. Bayesian consensus cladogram of 16S rDNA sequences from 58 cyanobacterial strains, based on GTR+G+I substitution model, with *Beggiatoa sp.* used as outgroup. Posterior probabilities (> 0.9) are shown in black at nodes and bootstrap values ($> 50\%$) in red. Posterior probabilities were calculated from 265,858 trees and bootstrap values from 500 re-samplings of the original data set. Colors define groups: yellow are single-celled cyanobacteria of section I; orange single-celled from section II; green are multicellular, undifferentiated cyanobacteria from section III; blue are multicellular and differentiated bacteria from section IV; and pink from section V. Sections as described by Castenholz 2001 [64]. AC, B, C, E and E1 denote phylogenetic clades described in the text.

cus, Oscillatoria, Plectonema). Clade C consists of *Pseudanabaena* species, *Arthronema gygaxiana* and *Phormidium mucicola* belonging to section III. *Gloeobacter violaceus* is placed closest to the outgroup. Several phylogenetic studies were conducted showing approximate agreement with the tree topology generated here [112, 115, 85, 116, 201, 141, 215, 118, 143, 119, 120]. To check the consistency of results from the maximum likelihood and Bayesian analysis to previous studies, we compare our results to the trees produced by Honda *et al.* [115], Turner *et al.* [85] who used 16S rDNA sequences, and Swingley *et al.* [119] who used a genomic approach.

The tree from Figure 2.2 in Honda *et al.* [115] shows overall strong congruences with our tree. The only exception is that in Honda *et al.* [115] "*Synechococcus elongatus* Toray" is placed separately between *Gloeobacter* and the rest of the cyanobacteria. We found that "*Synechococcus elongatus* Toray" (identical to *Thermosynechococ-*

cus elongatus BP1) is located within clade AC in our study and not next to *Gloeobacter violaceus*.

In Turner *et al.* [85], the major clades are congruent with those inferred in our study, but there are a few differences in the relationships among these clades. In that study, the analog of clade E1 is sister to clade AC, which is not the case in our consensus tree. Furthermore, *Synechococcus* C9 is grouped with *Synechococcus* P1, which might be due to long branch attraction. In our phylogenetic tree, *Synechococcus* C9 is grouped within clade AC, a relationship supported by high posterior probabilities and bootstrap values (1.0/99%). Clade C in our study is placed in the same position as in the tree from Turner *et al.* [85].

Swingley *et al.* [119], used a phylogenomic approach to investigate cyanobacterial relationships. Due to limited, biased genome data available at present, some clades present in our tree are missing in that study. Even so, the main clades retrieved in that study are mostly congruent with clades in our tree.

Monophyly of section V (the branching, differentiated cyanobacteria) shown in our tree agrees with Turner *et al.* [85] and other studies [141, 215]. Nonetheless it is possible that the monophyly of section V bacteria is due to limited taxon sampling, since polyphyly has been detected for section V in another study [142]. *Gloeobacter violaceus* is placed as the first diverging lineage in the phylogeny after the outgroup, as suggested by previous studies [115, 85, 116, 201, 215, 118, 143, 120]. Our phylogenetic reconstruction also confirms the placement of taxa belonging to section I and III throughout the tree [112, 115, 85, 116, 201, 141, 215, 118, 143, 120]. The finding that possibly none of the traditional morphological sections are monophyletic, clearly indicates that similar morphologies have been gained and lost several times during the evolutionary history of living cyanobacteria. Overall, the strong phylogenetic agreement between this and earlier studies confirms the suitability of the tree presented here for further analyses of morphological evolution.

Ancestral character state reconstruction

Our analysis indicates that multicellularity is a phylogenetically conservative character (p-value < 0.01). If the terminal taxa of the Bayesian consensus tree are randomly re-shuffled, a count through 1,000 re-shuffled trees gives an average of 20 transition steps. However an average of only nine parsimonious transitions was observed in a count through 10,000 randomly sampled trees of our ancestral character state reconstruction.

Results of the character state reconstruction using the AsymmMK model with transition rates estimated by Mesquite 2.71 [219] are displayed in Figure 2.5. Using maximum likelihood analysis, average frequencies of the characters were counted across 10,000 trees randomly sampled from the two Metropolis-coupled Markov Chain Monte Carlo (MC^3) searches of the Bayesian tree reconstruction.

In Figure 2.5, results of the ancestral character state reconstruction using maximum likelihood analysis with transition rates estimated by the program, are presented. Cyanobacteria share a unicellular ancestor, but multicellularity evolved early in the cyanobacterial lineage. We identified multicellular character states for three basic ancestors leading to clades E, AC and C in our tree. Together, these clades encompass the entirety of the morphological sections II, III, IV and V. Additionally character states were reconstructed using maximum likelihood analysis and fixed transition rates to analyze properties of the data set. Transition rates are presented in Table 2.3. Probabilities for character states at nodes 3, 4 and 5 were examined in detail (Table 2.4). A mul-

Tab. 2.3. Different Transition rates with whom ancestral character states were estimated. [1] *Asymmetrical Markov k-state 2 parameter model; rates estimated from the consensus tree.* [2] *Markov k-state 1 parameter model; rates estimated from the consensus tree.* [3] *F1-F6: Models using different fixed transition rates.* [4] *Reversible jump for model selection, using a hyper prior.* [5] *Forward rate describing changes to multicellularity.* [6] *Backward rate describing changes back to a unicellular state.*

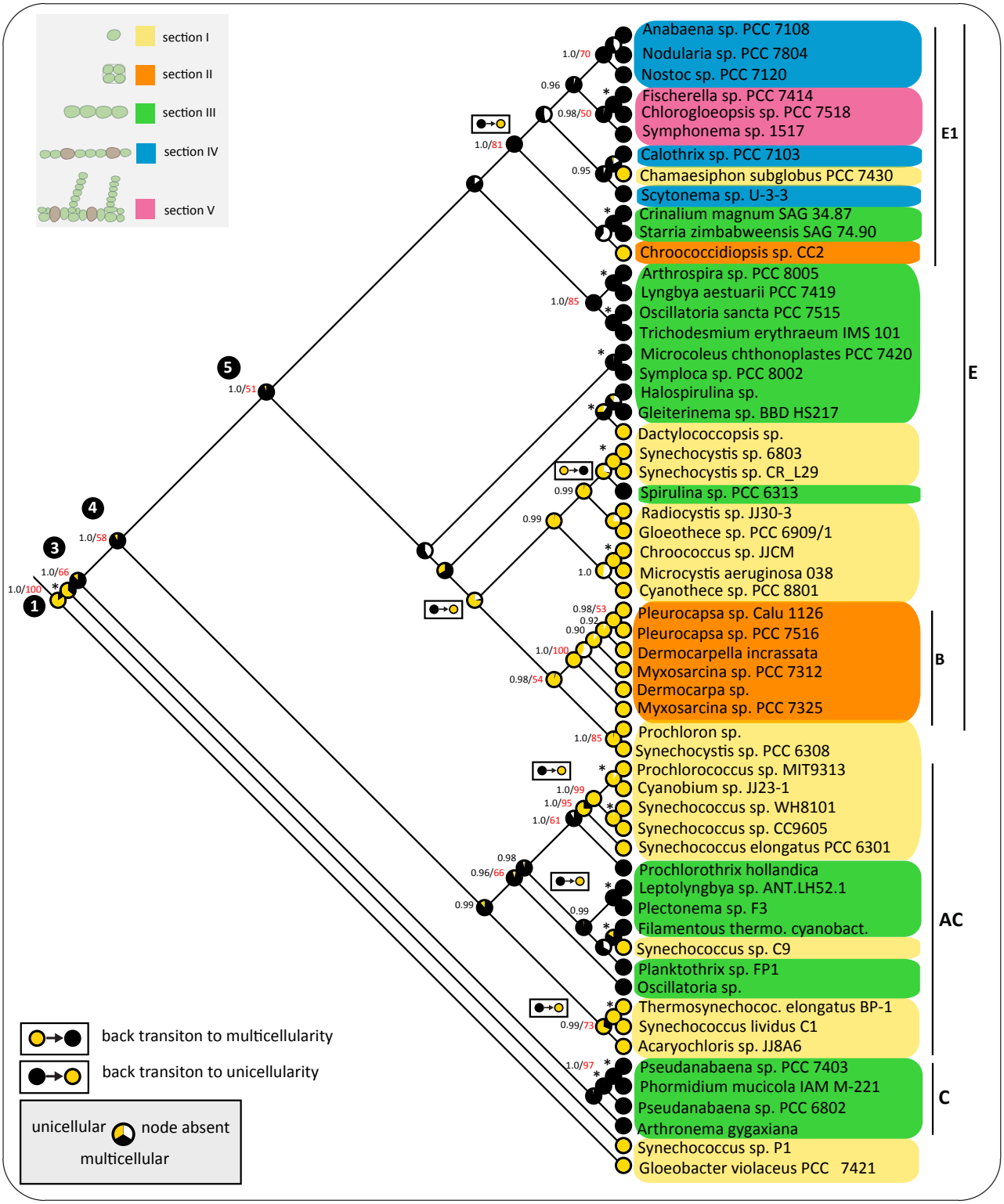
method	Maximum likelihood analysis								Bayesian analysis
rates	AsymmMK[1]	MK1[2]	F1[3]	F2	F3	F4	F5	F6	rjhp[4]
fw[5]	1.62	2.67	0.90	2.70	5.40	0.45	0.90	2.70	2.881
bw[6]	2.99	2.67	2.70	0.90	0.45	5.40	0.90	2.70	2.873

ticellular ancestry is very likely for these three nodes. For node 3 the relative probabilities of a multicellular ancestor range from 0.79 to 1.00, depending on the probability of the transition rates. For node 4 with varying transition rates, the relative probabilities of a multicellular ancestor range from 0.83 to 1.00. For node 5 the probabilities for multicellularity range from 0.90 to 1.00.

The maximum likelihood analysis is not contradicted by a Maximum Parsimony optimization (Table 2.4 and Additional File 2.7). Applying maximum parsimony as a reconstruction method, the uniquely best states were counted across 10,000 trees randomly sampled from the two (MC^3) runs of the Bayesian tree reconstruction. The relative probabilities for a multicellular ancestor at nodes 3, 4 and 5 are 0.68, 0.68 and 0.69, respectively. In contrast, the relative probabilities for a unicellular ancestor at nodes 3, 4 and 5 under parsimony reconstruction are 0.0013, 0.0014 and 0.0014 , respectively. Using Bayesian methods, a similar pattern is observed for these nodes. As an evolutionary model, BayesFactors revealed that a "hyperprior" approach with exponential prior distributions, whose means were sampled from a uniform distribution between 0 and 10 gave the best fit. Transition rates were estimated to be almost equal. Figure 2.6 displays the posterior probability distributions of character states at these three nodes as they were estimated over 10,000 randomly sampled trees. At nodes 3 and 5 posterior probabilities of a multicellular character state display values above 0.90 for most of the trees. At node 4 a multicellular state is more likely as well. Posterior probabilities at node 4 are above 0.75 for most of the trees.

At least five reversals to unicellularity occurred in the tree, three of them within clade AC. The first transition occurred on a branch which led to a group of thermophilic cyanobacteria: *Acharyochloris sp.*, *Synechococcus lividus* C1 and *Thermosynechococcus elongatus*. Posterior probabilities (PP) and bootstrap values (BV) for this group are 0.99/73%, whereas the sister group within AC is supported by 0.96/66% (PP/BV). The second tran-

Fig. 2.5 ***(on the next page).*** **Ancestral character state reconstruction using maximum likelihood.** Ancestral character state reconstruction with maximum likelihood analysis, using the "Asymmetrical Markov k-state 2 parameter"(AsymmMk) model implemented in Mesquite 2.71 [219]. Transition rates were estimated by the program (Table 2.3). Analysis was run over 10,000 randomly sampled trees from the Bayesian analysis and plotted on the Bayesian consensus tree. Possible states are unicellular (yellow) and multicellular (black). Relative likelihood probabilities for each character state are represented with a pie chart at nodes. The white part in the pie charts indicates the fraction of trees where the node was absent. Posterior probabilities (black) and bootstrap values (red) from the phylogenetic analyses are displayed at the nodes. Asterisks denote supported nodes for which posterior probabilities and bootstrap values are presented in Figure 2.4. At nodes 3, 4 and 5 a multicellular ancestry is very likely. Back mutations to unicellularity occur at least five times. A back mutation to multicellularity occurs at least once. Clades where transitions occurred are labelled.


section I
section II
section III
section IV
section V
Anabaena sp. PCC 7108
Nodularia sp. PCC 7804
Nostoc sp. PCC 7120
Fischerella sp. PCC 7414
Chlorogloeopsis sp. PCC 7518
Symphonema sp. 1517
Calothrix sp. PCC 7103
Chamaesiphon subglobus PCC 7430
Scytonema sp. U-3-3
Crinalium magnum SAG 34.87
Starria zimbabweensis SAG 74.90
Chroococcidiopsis sp. CC2
Arthrospira sp. PCC 8005
Lyngbya aestuarii PCC 7419
Oscillatoria sancta PCC 7515
Trichodesmium erythraeum IMS 101
Microcoleus chthonoplastes PCC 7420
Symploca sp. PCC 8002
Halospirulina sp.
Gleiterinema sp. BBD HS217
Dactylococcopsis sp.
Synechocystis sp. 6803
Synechocystis sp. CR_L29
Spirulina sp. PCC 6313
Radiocystis sp. JJ30-3
Gloeothece sp. PCC 6909/1
Chroococcus sp. JJCM
Microcystis aeruginosa 038
Cyanothece sp. PCC 8801
Pleurocapsa sp. Calu 1126
Pleurocapsa sp. PCC 7516
Dermocarpella incrassata
Myxosarcina sp. PCC 7312
Dermocarpa sp.
Myxosarcina sp. PCC 7325
Prochloron sp.
Synechocystis sp. PCC 6308
Prochlorococcus sp. MIT9313
Cyanobium sp. JJ23-1
Synechococcus sp. WH8101
Synechococcus sp. CC9605
Synechococcus elongatus PCC 6301
Prochlorothrix hollandica
Leptolyngbya sp. ANT.LH52.1
Plectonema sp. F3
Filamentous thermo. cyanobact.
Synechococcus sp. C9
Planktothrix sp. FP1
Oscillatoria sp.
Thermosynechococ. elongatus BP-1
Synechococcus lividus C1
Acaryochloris sp. JJ8A6
Pseudanabaena sp. PCC 7403
Phormidium mucicola IAM M-221
Pseudanabaena sp. PCC 6802
Arthronema gygaxiana
Synechococcus sp. P1
Gloeobacter violaceus PCC 7421
E1
E
B
AC
C
back transiton to multicellularity
back transiton to unicellularity
unicellular
node absent
multicellular

Tab. 2.4. Ancestral character states of nodes 3, 4 and 5 using different transition rates and methods. [1] *Maximum likelihood: Average frequencies across trees were calculated.* [2] *Maximum parsimony: Uniquely best states across trees were counted.* [3] *Bayesian analysis: model parameters estimated based on the data.*

			node 3		node 4		node 5	
method	model		state1	state0	state1	state0	state1	state0
ML[1]	AsymmMK	estimated[3]	0.88	0.12	0.91	0.08	0.95	0.05
		F1	0.96	0.04	0.98	0.02	0.99	0.01
		F2	0.87	0.12	0.91	0.09	0.94	0.06
		F3	1.00	0.00	1.00	0.00	1.00	0.00
		F4	0.88	0.12	0.92	0.08	0.95	0.05
	MK1	estimated[3]	0.79	0.21	0.83	0.17	0.90	0.10
		F5	0.88	0.12	0.90	0.10	0.93	0.07
		F6	0.79	0.21	0.83	0.17	0.90	0.10
MP[2]			0.6805	0.0013	0.6799	0.0014	0.6871	0.0014
BA[3]	rjhp		0.915	0.0851	0.817	0.183	0.902	0.0980

sition within clade AC led also to a thermophilic cyanobacterium *Synechococcus* C9. Sister relation of this species to a filamentous thermophilic cyanobacterium is supported by 1.0/99% (PP/BV). The last transition in clade AC occurred within the group including the marine pico-phytoplankton genera *Synechococcus* and *Prochlorococcus*. The filamentous *Prochlorothrix hollandica* is supposed to be the closest relative to the group that includes marine pico-phytoplankton, supported by 1.0/61% (PP/BV). Clade AC has a PP of 0.99, while its BV is below 50%. Although bootstrap support is below 70% for clade AC and some groups within it, posterior probabilities show a very high support (> 0.95). Simulation studies have shown that posterior probabilities approach the actual probability of a clade [220, 221, 222]. Bootstrapping tends to underestimate the actual probability of a true clade. Although, posterior probabilities tend to be erroneous if the model of evolution is underparameterized, overparameterization has only a minor effect on the posterior probabilities. Therefore, using a complex model of evolution, such as the "general time reversible with *gamma* distributed rate variation"(GTR+G), is recommended [221, 222]. We used the GTR+G+I model for our analysis, and assume that nodes with a PP higher than 0.95 are reliable.

It is very likely that at least one additional reversal to unicellularity occurred in clade E1, but phylogenetic support is not high enough to locate the exact position of this transition. Similarly, support for the nodes where the other transition to multicellularity within clade E occurred is missing. The exact locations of reversals within clade E therefore are not certain and a scenario where multiple reversals occurred cannot be excluded. In clade E, there is also a reversal to multicellularity observed in *Spirulina sp.* PCC 6313. The location of this transition is supported by posterior probabilities of 0.99 at two ancestral nodes.

Stucken *et al.* [223] compared gene sets of multicellular cyanobacteria and found that at least 10 genes are essential for the formation of filaments. Besides genes previously thought to be correlated with heterocyst formation (*hetR*, *patU3* and *hetZ*) they found seven genes coding for hypothetical proteins. The species they compare are all located within clade E in our tree, most of them being differentiated. Unfortunately no genomes from multicellular species in more basal clades are available at present. But genome projects of *Phormidium sp.* ISC 31 and *Plectonema sp.* ISC 33 are presently being conducted (*http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi*).

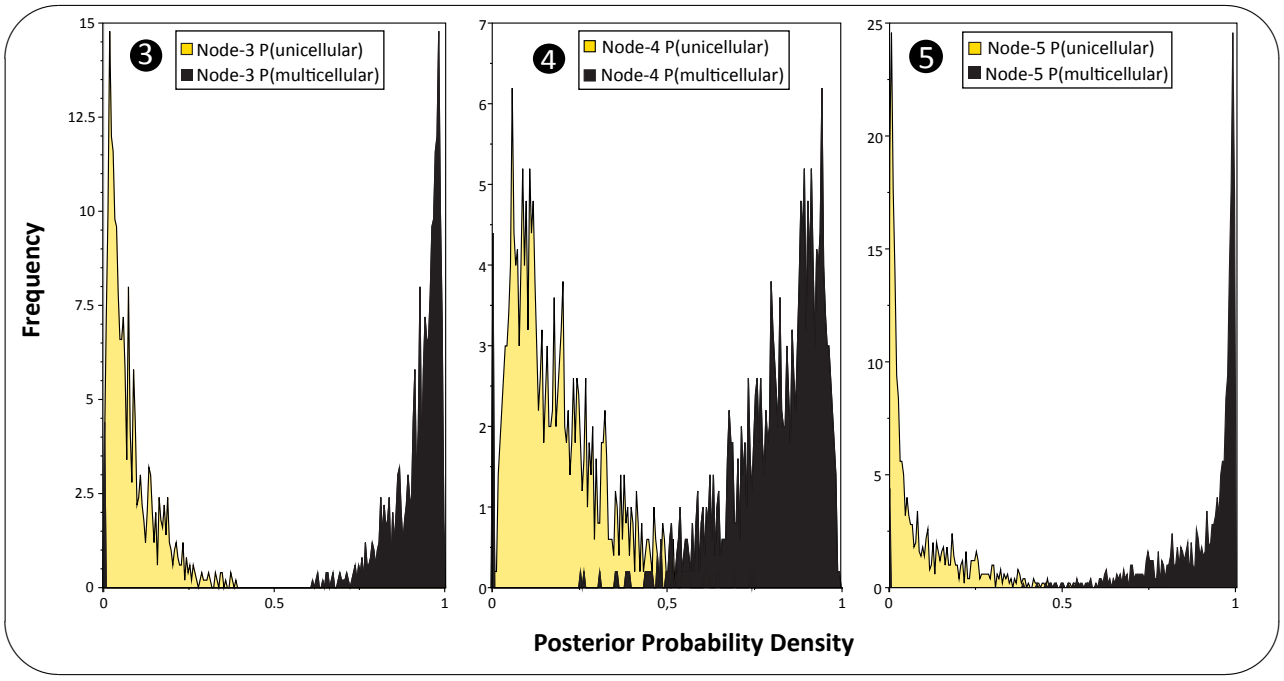


Fig. 2.6. Ancestral character states of nodes 3, 4 and 5 using Bayesian analysis. Posterior probability distribution for a unicellular character state (yellow) and a multicellular character state (black) at nodes 3, 4 and 5 from 10,000 Bayesian trees. 2x 5,000 trees were randomly sampled from 2 MC^3-searches. Analysis was performed using BayesTraits. Posterior distributions were derived from reversible jump MCMC-search of 30 million iterations using a hyperprior approach. The probability of a multicellular ancestry is shifted towards 1 for each of the three nodes.

If these species turn out to group with *Phormidium mucicola* IAM M-221 and *Plectonema sp.* F3 from the basal clades C and AC in our study, this could provide important information on the original metabolic pathways in ancient multicellular cyanobacteria and on possible advantages of multicellularity.

The majority of cyanobacteria living today are described as successful ecological generalists growing under diverse conditions [106]. Our analysis indicates that this diverse range of cyanobacterial morphotypes found in various habitats today —whether multicellular or unicellular— has evolved from multicellular ancestors.

Gaining and losing multicellularity

In eukaryotes, simple multicellular forms build the foundation for the evolution of complex multicellular organisms. Although complex multicellularity exhibiting more than three cell types is presumably missing in prokaryotes, bacteria invented simple multicellular forms possibly more than 1.5 billion years earlier than eukaryotes [224, 108, 196, 197]. Multicellularity has been described as one of several major transitions that occurred in the history of life. These transitions between different units of selection [225] resulted in changes in the organizational confines of the individual. Maynard Smith and Szathmary [226](1995, p.6) summarize eight major transitions in the evolution of life after which, "entities that were capable of independent replication before the transition can replicate only as part of a larger whole after it". These transitions can create new units of selection at a higher level of complexity [227]. Origin of chromosomes, origin of the eukaryotic cell, origin of multicellular organisms and the origin of eusocial communities are some major transitions that redefine the

degree of individuality [225, 226, 228, 229]. Some transitions are thought to be unique, such as the evolution of meiosis or the evolution of the genetic code. Other major transitions occurred several times independently, such as the evolution of eusociality [230, 231] and multicellularity [225, 185, 232, 233, 234]. There is a tendency to assume that these transitions occur in a progression that leads to an increase in complexity. However, it seems that in cyanobacteria this is not the case. Anatomical complexity has been lost during their evolution several times (Figure 2.5). In a similar fashion, a complex character such as eusociality has been lost several times in halictid bees [231, 235]. Conversely the phylogeny indicates that multicellularity re-evolved in *Spirulina*. Regaining complex characters has been observed in other studies as well [236, 237, 238]. Nonetheless, some studies state that re-evolution of a complex character after a previous loss is not possible [239, 240]. Such studies argue that according to 'Dollo's law', a loss of complexity is irreversible [241], a statement that is not supported in the cyanobacterial case. Repeated transitions in either direction are possible.

Prokaryotic fossil record before the "Great Oxygenation Event": Evidence for multicellular cyanobacteria?

Various claims for life during the early Archean Eon, more than 3.00 billion years ago exist. Most of them from two regions: the Berberton Greenstone Belt, South Africa (around 3.20-3.50 billion years old) and the Pilbara Craton, Western Australia (around 2.90-3.60 billion years old). For some of these "fossils" a biological origin is questioned [105, 242, 197], but for others biogenicity is very likely [195, 243, 196, 244, 245, 246, 197]. These candidates for early life have clear age constraints and there is no non-biological explanation for these structures. The ages and possible metabolic features of seven fossils of proposed biological origin are plotted in Figure 2.7 (1-7) [195, 243, 196, 244, 245, 246, 197]. Some of these fossils are assumed to have been photosynthetic and mat builders, characteristics that can be identified in cyanobacteria as well. One of the oldest fossils recorded, 3.45 billion year old prokaryotic remains found in the Panorama Formation, East Pilbara Craton, Western Australia exhibit a filamentous morphotype and possibly carried out anoxygenic photosynthesis [196, 197]. Some late Archean fossils show an oscillatorian or chroococcacean morphotype (Figure 2.7, 8-9). 2.52 and 2.56 billion year old oscillatorian-like fossils [247, 248, 108] could possibly represent close relatives of cyanobacterial ancestors. 2.72 billion year old filamentous bacteria [108] could potentially represent one of the first multicellular cyanobacteria detected. For single celled forms, 2.56 billion year old unicellular fossils [249, 250, 251, 248] could likely represent chroococcacean fossils, relatives of ancestral *Gloeobacter violaceus* or *Synechococcus sp.* P1 (Figure 2.7).

The first conclusive cyanobacterial fossils from all five sections have been reported from around 2.15 billion year old rocks. In 1976, Hofmann described Microfossils from stromatolitic dolomite stones in the Kasegalik and McLeary Formations of the Belcher Supergroup in Hudson Bay, Northern Canada. Among these fossils are *Halythrix* which seems to belong to the order Oscillatoriales (section III), *Eosynechococcus* and *Entophysalis* both presumably order Chroococcales (section I) and *Myxococcoides* fossils (section II). In 1997 similar fossils were described by Amard and Bertrand-Safarti in paleoproterozoic cherty stromatolites from the "Formation C (FC)" of the Franceville Group in Gabon, dating back 2.00 billion years. They also characterized chroococ-

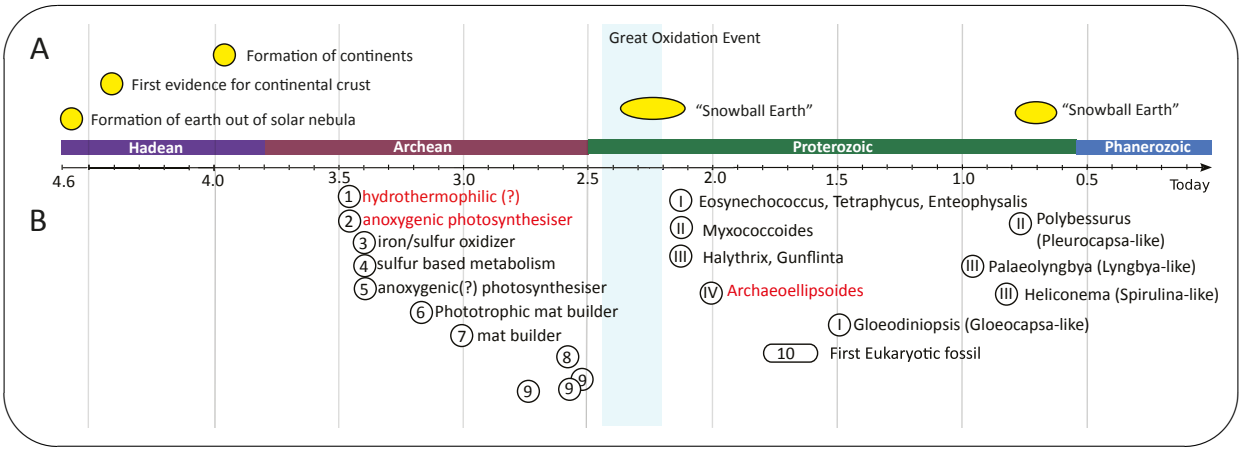


Fig. 2.7. Timeline with prokaryotic fossil record. Timeline with geological events (A) and prokaryotic fossil record (B). (A) Formation of Earth [252], first evidence of continental crust [253], formation of continents [252], and glaciation events described in the Snowball Earth hypothesis [254]. (B) The oldest conclusive cyanobacterial fossils are found in around 2.15 billion year old rocks. 1-7: Fossils from the Archean Eon [195, 243, 196, 244, 245, 246, 197]. 8: chroococcacean fossils [108]; 9: oscillatorian fossils[108]. I-V: cyanobacterial fossils [103, 106, 123]. 10: eukaryotic fossils [224].

calean fossils, particularly *Eosynechococcus* and *Tetraphycus*, and filamentous bacteria (*Gunflinta*) which could likely resemble cyanobacteria and *Myxococcoides* fossils. Furthermore, large microfossils (so called *Archaeoellipsoides elongatus*), with akinetes similar to the ones from Anabaena-like species were found [123, 143]. Akinetes are resting cells which are only present in differentiated cyanobacteria from sections IV and V. As it has been confirmed in several studies, sections IV and V share a most recent common ancestor [85, 141, 143]. Therefore these fossil akinetes document the existence of differentiated cyanobacteria 2.00 billion years ago. Given that differentiation in cyanobacteria is evolutionary stable only in a multicellular setting [255], this again supports the notion that multicellular species belonging to the cyanobacteria must have existed earlier than 2.0 billion years ago.

Several studies have assessed prokaryotic history using phylogenetic dating methods [211, 213]. In these studies the origin of cyanobacteria has been estimated around the time of the "Great Oxygenation Event" of 2.20-2.45 billion years ago [110, 111]. Other studies have reported elevations of oxygen levels before the great rise of atmospheric oxygen [256, 111]. Using small and large ribosomal subunit sequences, Blank and Sanchez-Baracaldo [120] estimated the origin of cyanobacteria between 2.7 and 3.1 billion years ago. They also try to address the evolution of cyanobacterial traits and assess that multicellular cyanobacteria did not originate before 2.29-2.49 billion years ago. In the study of Blank and Sanchez-Baracaldo [120], a smaller set of cyanobacterial taxa was used, with some basal multicellular species that are present in clade C of our analysis missing. These taxa could have an essential effect on the timing of the first multicellular cyanobacteria. To resolve this issue further dating analyses would be needed. Clearly, as Blank and Sanchez-Baracaldo point out, for such analyses to ultimately resolve the cyanobacterial history, a larger number of cyanobacterial genome data would be needed to represent all the morphological and genetic diversity within this phylum.

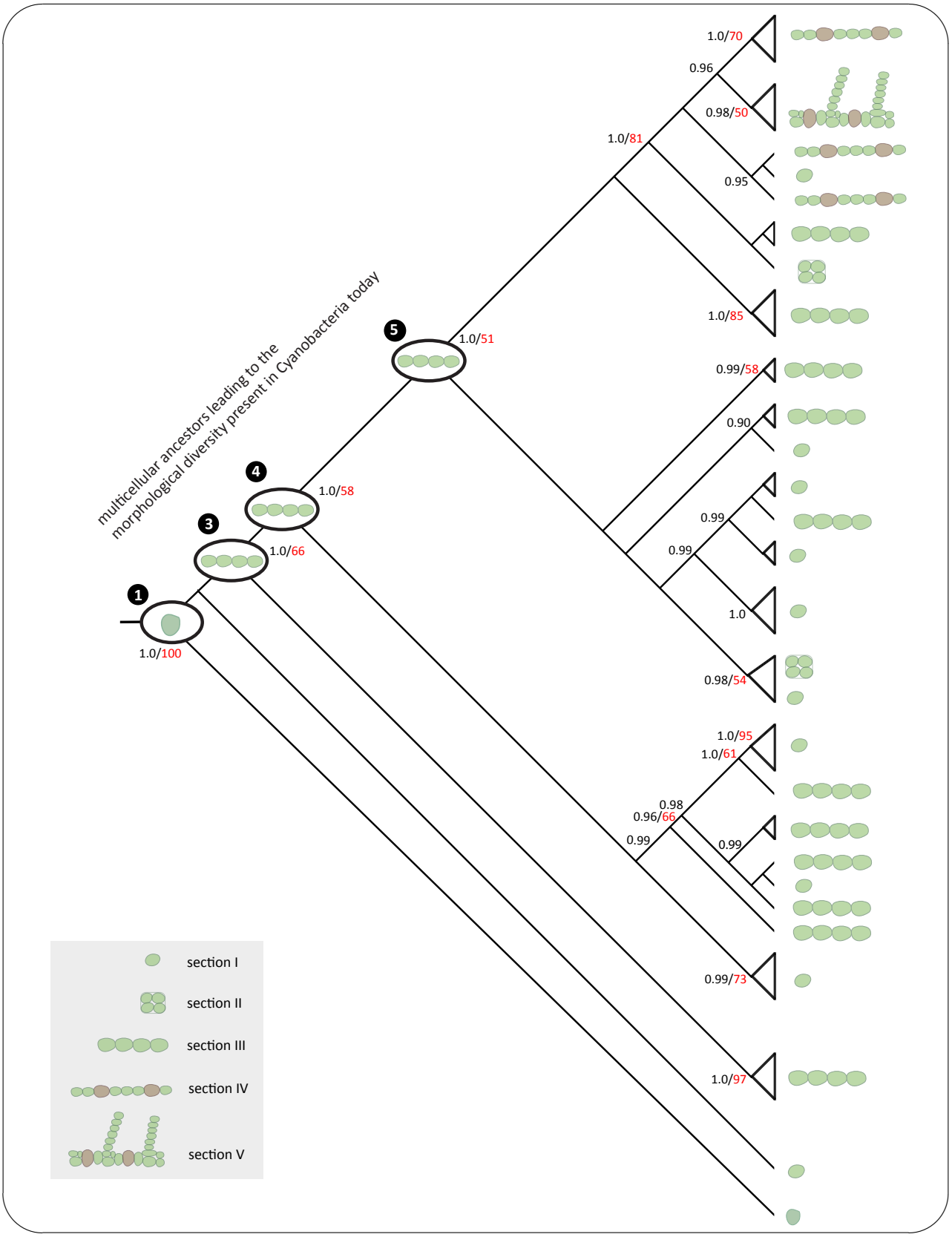


Fig. 2.8. Schematic illustration of cyanobacterial evolution. Numbers at the nodes indicate Bayesian posterior probabilities (black) and bootstrap values (red) from the phylogenetic analyses. The most recent common ancestor of all cyanobacteria is optimized to have been unicellular. All cyanobacteria derive from a unicellular most recent common ancestor (node 1). The lineage leading to *Gloeobacter violaceus* diverges very early from the remaining cyanobacteria. Most major clades of cyanobacteria derive from multicellular ancestors (nodes 3-5).

Conclusions

Cyanobacteria, photosynthetic prokaryotes, are one of the oldest phyla still alive on this planet. Approximately 2.20-2.45 billion years ago cyanobacteria raised the atmospheric oxygen level and established the basis for the evolution of aerobic respiration [181, 110, 182, 143, 183, 184]. They introduced a dramatic change in the Earth's atmosphere, which might have created possibilities for more complex lifeforms to evolve. Considering the importance of cyanobacteria for the evolution of life, it seems unfortunate that data sets for a representative phylogenomic analysis are not yet available. A coordinated perspective between research groups and a diversified taxon sampling strategy for genome projects would offer the possibility for more comprehensive studies on cyanobacterial evolution. By presenting results obtained from 16S rDNA data analysis here, we hope to boost interest for more extensive genomic studies in this phylum. Phylogenomic approaches would help to further investigate some of the results in the present work.

Multicellular prokaryotic fossils from the Archean Eon are documented [196, 197], and fossil data can support the possibility of multicellular cyanobacteria in the Archean Eon [247, 249, 248, 108]. Furthermore, studies describe smaller accumulations of oxygen levels around 2.8 to 2.6 billion years ago [111] and around 2.5 billion years ago [256]. Therefore multicellular cyanobacteria could have evolved before the rise of oxygen in the atmosphere. The "Great Oxygenation Event", also referred to as "oxygen crisis", could presumably have marked one of the first mass extinction events during Earth's history. New habitats developing around 2.32 billion years ago, due to a dramatic change of Earth's atmosphere could have triggered cyanobacteria to evolve the variety of morphotypes preserved until today.

In terms of cell types, cyanobacteria reached their maximum morphological complexity around 2.00 billion years ago[257]. By the time eukaryotes evolved, cyanobacteria already exhibited the full range of their morphological diversity. Due to slow evolutionary rates in cyanobacteria, which have been described as "hypobradytelic" [106, 217, 218], extant cyanobacteria that appear to exhibit the same morphotype as in the Precambrian Eon [258] are reminiscent of the idea of "living fossils". However, one should consider the possibility that what may appear as morphological stasis may be due to developmental constraints at the phylum level. Cyanobacteria apparently reached their maximum complexity early in Earth history, but instead of morphological stasis at the species level, our results suggest that they subsequently changed morphotypes several times during their evolution. This allowed for the exploration of diverse morphotypes within their developmental constraints, including the loss and regaining of multicellular growth forms.

Figure 2.8 summarizes the morphological evolution of the cyanobacteria inferred in this study. All extant cyanobacteria share a most recent common ancestor that was unicellular. Single-celled species at the base of the tree do not seem to have changed much in their morphology and are possibly comparable to ancient cyanobacteria. Aside from *Gloeobacter violaceus* and *Synechococcus* P1, which diverged very early, all cyanobacteria living today share multicellular ancestors. Although complex multicellularity is missing in prokaryotes, these simple multicellular forms have evolved several hundred million years before the appearance of eukaryotes, whose fossil record dates back to 1.8-1.3 billion years ago [224]. In agreement with various proposed selective advantages that multicellular growth could confer [259, 260, 261, 262], the results presented here indicate that the

early origin of multicellularity played a key role in the evolutionary radiation that has led to the majority of extant cyanobacteria on the planet.

Methods

Taxon sampling

A total of 2,065 16S rRNA gene sequences from the phylum cyanobacteria were downloaded from GenBank. Unidentified and uncultured species were excluded. With this large dataset phylogenetic reconstructions were conducted as described in the next section. Aside from cyanobacteria, the dataset included six chloroplast sequences and six eubacterial sequences: *Beggiatoa sp.*, *Thiobacillus prosperus*, *Agrobacterium tumefaciens*, *Chlorobium sp.*, *Candidatus Chlorothrix halophila* and *Escherichia coli HS* (Additional File 2.1).

From this large tree a subset of 58 cyanobacterial sequences were selected for further analyses. Accession numbers are provided in Table 2.1. Species from all five sections described by Castenholz *et al.* [64] were included. Taxa were chosen to represent a 1:1 ratio of unicellular and multicellular species. The final data set contained 22 single-celled taxa from section I, 7 single-celled taxa from section II, 21 multicellular taxa from section III, 5 multicellular, differentiated taxa from section IV and 3 differentiated, branching taxa from section V as described by Castenholz *et al.* [64].

An outgroup for further analyses was chosen from a set of eubacterial, non cyanobacterial species whose 16S rRNA gene sequences were downloaded from GenBank. Species were sampled to cover a wide range of different phyla. Aside a set of species from phyla represented in the "tree of life" [212], species from additional phyla as described on NCBI (http://www.ncbi.nlm.nih.gov/Taxonomy/ - Taxonomy Browser: Bacteria) were selected for analyses (Table 2.2).

Phylogenetic analyses

Phylogenetic analyses of all identified cyanobacteria

The 2,065 16S rRNA gene sequences were aligned using the software MAFFT [263] via Cipres Portal [264]. The alignment was corrected manually using BioEdit v7.0.5 [265]. Poorly aligned and duplicated sequences were excluded from the alignment. From the remaining 1,254 sequences (1235 characters) a phylogenetic tree was reconstructed running 10 maximum likelihood analyses as implemented in RAxML v7.0.4 [266]. GTR + G + I (General time reversible model, G: Gamma correction, I: proportion of invariable sites) [267, 268] was used as an evolutionary substitution model. Bootstrap values were calculated from 100 re-samplings of the dataset and plotted on the best maximum likelihood tree using RAxML v7.0.4. The resultant tree (Figure 2.1; Additional File 2.2: newick format) was visualised in FigTree v1.3.1 ($http://tree.bio.ed.ac.uk/software/figtree/$) and graphically edited with Adobe Illustrator CS2 ($http://www.adobe.com/products/illustrator/$).

Phylogenetic analyses to identify an outgroup

To test different outgroups, phylogenetic trees were reconstructed using all sampled non-cyanobacterial species (Table 2.2) plus five representative species from the cyanobacterial phylum (Table 2.1). Sequences were aligned using Clustal-X with default settings [175] and corrected manually. The trees were built using maximum likelihood and Bayesian inference, with and without an outgroup from the kingdom archaea. 50 separate maximum likelihood searches were conducted using RAxML v7.0.4 software [266], from which the tree with the best log-likelihood was chosen. Bootstrap support for each tree was gathered from 100 re-samplings. Bayesian analyses were conducted with MRBAYES 3.1 [176] using a GTR + G + I evolutionary model with substitution rates, base frequencies, invariable sites and the shape parameter of the gamma distribution estimated by the program. Two Metropolis-coupled Markov Chain Monte Carlo (MC^3) searches with four chains, three heated and a cold one, were run. The analyses started with a random tree and was run for 5,000,000 generations. Trees and parameters were sampled every 100th generation. The trees were checked to show a standard deviation of split frequencies below 0.05. The first 3,000,000 generations were excluded as the burn-in.

Additionally phylogenetic analyses were conducted with Bayesian inference, using each of the 22 eubacterial species separately with the sampled cyanobacterial subset (58 taxa). Alignments were built using Clustal-X software with default settings [175] and corrected manually. For each phylogenetic analysis two (MC^3) searches were run for 10,000,000 generations using MRBAYES 3.1 [176]. Trees and parameters were sampled every 100th generation. The first 3,000,000 generations being excluded as a burn-in, assuring that the standard deviation of split frequencies were below 0.05 and log-likelihoods of the trees had reached stationarity. Results were compared and *Beggiatoa sp.* was chosen as an outgroup for further analyses.

Phylogenetic analyses of a cyanobacterial subset

Sequence alignments of the 16S rRNA gene sequences from the cyanobacterial subset and *Beggiatoa sp.* (59 taxa, 1166 characters) were carried out using Clustal-X with default settings [175] and corrected manually. Whether the cyanobacterial alignment (excluding the outgroup) was substitutionally saturated was tested using the program DAMBE [**?** 269]. The information-entropy based index of substitutional saturation [270] was used to analyze our alignment of 16S rRNA gene sequences. The test performs only on a maximum of 32 species. Therefore we sampled from our phylogeny 32 representative sequences that span the whole tree, and performed the test introduced by Xia et al.[270](Table 2.1 and Additional File 2.5).

Phylogenetic reconstruction was carried out using Bayesian analysis and maximum likelihood. Maximum likelihood analysis was performed using GARLI 0.96 [271] and Bayesian analysis was conducted with MRBAYES 3.1 [176]. The evolutionary model of nucleotide substitution that best fitted the data was obtained by using the Akaike Information Criterion as implemented in Modeltest 3.5 [272]. The selected model was GTR + G + I. Substitution rates, base frequencies, invariable sites and the shape parameter of the gamma distribution were estimated by the program. Fifty maximum likelihood searches were performed. Bootstrap values were calculated from 500 re-samplings of the data set. The bootstrap values were plotted on the best ML-tree using the program SumTrees [273].

Bayesian analysis was conducted running two (MC^3) searches, each with four chains, one cold and three heated. Starting with a random tree, analyses were run for 16,616,000 generations each, with trees being sampled every 100th generation. The trees were checked for convergence of parameters (standard deviation of split frequencies below 0.01, effective sample sizes above 200, potential scale reduction factor equal to 1.0) using Tracer v1.4.1 [177] and the program AWTY [274]. Burn-in was set to 3,323,200 generations each, corresponding to the first 20% of the analyses. The average standard deviation of split frequencies was below 0.01 for the remaining 132,929 trees of each run, indicating that steady state of the log-likelihoods was reached.

Ancestral character state reconstruction

Character state reconstructions were performed using maximum parsimony (MP; Additional File 2.7) and maximum likelihood criteria as implemented in Mesquite 2.71 [219]. 5,000 trees from each MC^3 run were randomly chosen from the post burn-in Bayesian sample and combined. Discrete characters were coded into multicellular or unicellular states. The results over 10,000 Bayesian trees were summarized and displayed on the consensus tree of the Bayesian analysis. For maximum likelihood estimates, both the "Markov k-state 1 parameter model" (MK1 model) and "Asymmetrical Markov k-state 2 parameter model" (AsymmMK model) were applied. Rate of change is the only parameter in the MK1 model. The AsymmMK model exhibits two parameters, describing the forward and backward transitions between states. Phylogenetic conservativeness of multicellularity was tested by comparing the observed distribution of parsimony steps across 10,000 randomly chosen trees from the Bayesian analysis against the distribution from 1,000 trees modified from the Bayesian consensus by randomly shuffling the terminal taxa, while keeping the relative proportion of states unaltered. The root was assumed to be at equilibrium. Transition rates for the MK1 and AsymmMK model were estimated by the program. Rates for the latter models presented in Table 2.3 were estimated for the consensus tree. To explore properties of the data set, character states were additionally reconstructed with manually fixed transition rates (F1-F6; Table 2.3). The state of the outgroup was excluded from the analyses to avoid biased inferences within the ingroup.

The character states of nodes 3, 4 and 5 of the Bayesian consensus tree were additionally estimated using a reversible jump MCMC search as implemented in BayesTraits [275]. MCMC was run for 30 million iterations, and a burnin set to 50,000. The analysis was run several times with parameters of the evolutionary model being chosen from different prior distributions. In order to determine which model offered the best fitting priors, models were tested using Bayes Factors. A hyperprior approach with mean-values of the exponential priors derived from a uniform distribution between 0 to 10 was determined to fit best the data. The results of the analysis were visualized in Tracer v1.5 [177].

Author contributions

BES and HCB conceived the study; BES gathered data and conducted analyses; BES, HCB, AA designed research and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Elena Conti and Brian R. Moore for helpful comments on an earlier version of our manuscript. Furthermore, we would like to thank Marco Bernasconi whose comments on the final version were of great help, and Jurriaan de Vos for help with the software BayesTraits.

Corrections published along with the article:

—Methods—

1)

Taxon Sampling
Sentence 1: "A total of 2,065 16S rRNA gene sequences from the phylum cyanobacteria were downloaded from GenBank.", is not correct.

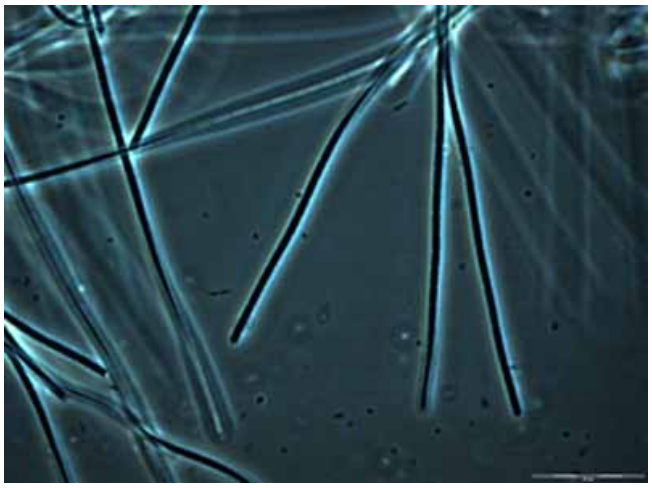
Instead it should read: "A total of 2,064 16S rRNA gene sequences from the phylum cyanobacteria were downloaded from GenBank."

2)

Phylogenetic analyses
Phylogenetic analyses of all identified cyanobacteria
Sentence 4: "From the remaining 1,254 sequences (1235 characters) a phylogenetic tree was reconstructed running 10 maximum likelihood analyses as implemented in RAxML v7.0.4 [104].", is not correct.

Instead it should read: "From the remaining 1,220 sequences (1261 characters) a phylogenetic tree was reconstructed running 10 maximum likelihood analyses as implemented in RAxML v7.0.4 [104].

CHAPTER III



Evolution of cyanobacterial morphotypes: taxa required for improved phylogenomic approaches

Authors:

Bettina E. SCHIRRMEISTER, Maria ANISIMOVA, Alexandre ANTONELLI and Homayoun C. BAGHERI

Schirrmeister et al. (2011) Communicative and Integrative Biology 4(4):424-427 - cover article

Addendum to: Schirrmeister BE, Antonelli A, Bagheri HC. The origin of multicellularity in cyanobacteria. BMC Evol. Biol. 2011;11:45; PMID: 21320320; DOI: 10.1186/1471-2148-11-45.

Within prokaryotes cyanobacteria represent one of the oldest and morphologically most diverse phyla on Earth. The rise of oxygen levels in the atmosphere 2.32-2.45 billion years ago is assigned to the photosynthetic activity of ancestors from this phylum. Subsequently cyanobacteria were able to adapt to various habitats evolving a comprehensive set of different morphotypes. In a recent study we showed that this evolution is not a gradual transition from simple unicellular to more complex multicellular forms as often assumed. Instead complexity was lost several times and regained at least once. An understanding of the genetic basis of these transitions would be further strengthened by phylogenomic approaches. However, considering that new methods for phylogenomic analyses are emerging, it is unfortunate that genomes available today are comprised of an unbalanced sampling of taxa. We propose avenues to remedy this by identifying taxa that would improve the representation of phylogenetic diversity in this phylum.

Cyanobacteria are a group of prokaryotes that receive their energy through oxygenic photosynthesis. The rise of oxygen levels in Earth's atmosphere 2.32-2.45 billion years ago, often referred to as the "Great Oxygenation Event" [110], provides the first indirect evidence for the presence of cyanobacteria. Based on their morphology, cyanobacteria represent one of the most diverse prokaryotic phyla. They have been divided into five different morphological sections [64]. Sections I and II comprise unicellular species, whereas sections III, IV and V consist of multicellular species. Sections IV and V evolved terminal differentiation and therefore represent the most morphologically complex cyanobacteria. Fossil records of cyanobacteria are well established and can be found in 2.0 billion year old rock formations [123]. In a recently published study, we traced the origin of multicellularity in cyanobacteria, and found a pattern of frequent morphological transitions [159]. A fast heuristic maximum likelihood approach made it possible to reconstruct a phylogeny of cyanobacteria for over a thousand taxa based on 16S rRNA sequences. A smaller subset of cyanobacteria representing full morphological and genetic diversity was used for the maximum likelihood and Bayesian reconstruction of ancestral character states. These analyses demonstrated that all cyanobacteria share a unicellular most recent common ancestor. However, multicellularity evolved very early in the cyanobacterial lineage. Evolution and divergence from the ancient multicellular lineage led to the richness of morphotypes manifested by cyanobacteria today. In the course of the cyanobacterial diversification process, complexity in the form of multicellular growth was lost at least five times. It was also regained at least once. Furthermore, by comparing the results to the prokaryotic fossil record, it was possible to elucidate that multicellularity evolved in cyanobacteria earlier than 2.0 billion years ago, possibly prior to the "Great Oxygenation Event". Several multicellular prokaryotic fossils that are suggestive of cyanobacterial morphotypes have been found in rock formations from the Archean Eon, which corroborates our conclusions[197].

For this addendum, we reconstructed two cyanobacterial phylogenies inferred with the general substitution model (GTR+G+I) from 16S rRNA sequences, and emphasize the position of fully sequenced genomes in the sample. Figure 3.1 displays a phylogenetic tree reconstructed from 1,261 cyanobacterial taxa (1,206 sites) using maximum likelihood as implemented in RAxML v7.0.4 [266]. Branch supports represent bootstrap values calculated from 100 re-samplings of the dataset. Figure 3.2 shows the phylogenetic tree built with a Bayesian approach as implemented in MrBayes v3.1 [176]. Two Metropolis-coupled MCMC runs were performed over 20,000,000 generations, sampling trees and parameters every 1,000th generation. Convergence of the parameters was checked using AWTY [274] and the first 50% of trees (namely 10,000) were discarded as burn-in. The trees were partitioned in groups (A1-A5; AC1; B1,B2; C1-C4; D1,D2) as previously described [159]. Upon examining the figures, an over-representation of genome data from unicellular cyanobacterial taxa is striking. Out of 41 fully sequenced cyanobacterial strains, 35 are single celled. All fully sequenced unicellular taxa belong to morphological section I, and are located in groups A1-A5 in our phylogeny. The six multicellular taxa for which full genome data are available belong to morphological sections III and IV, which are located in groups D1 and C2 (Fig.1 and 2). Although it seems that the fully sequenced strains belong to taxa-rich phylogenetic groups, this pattern could result from a bias in sampling or cultivation success. Therefore the apparent

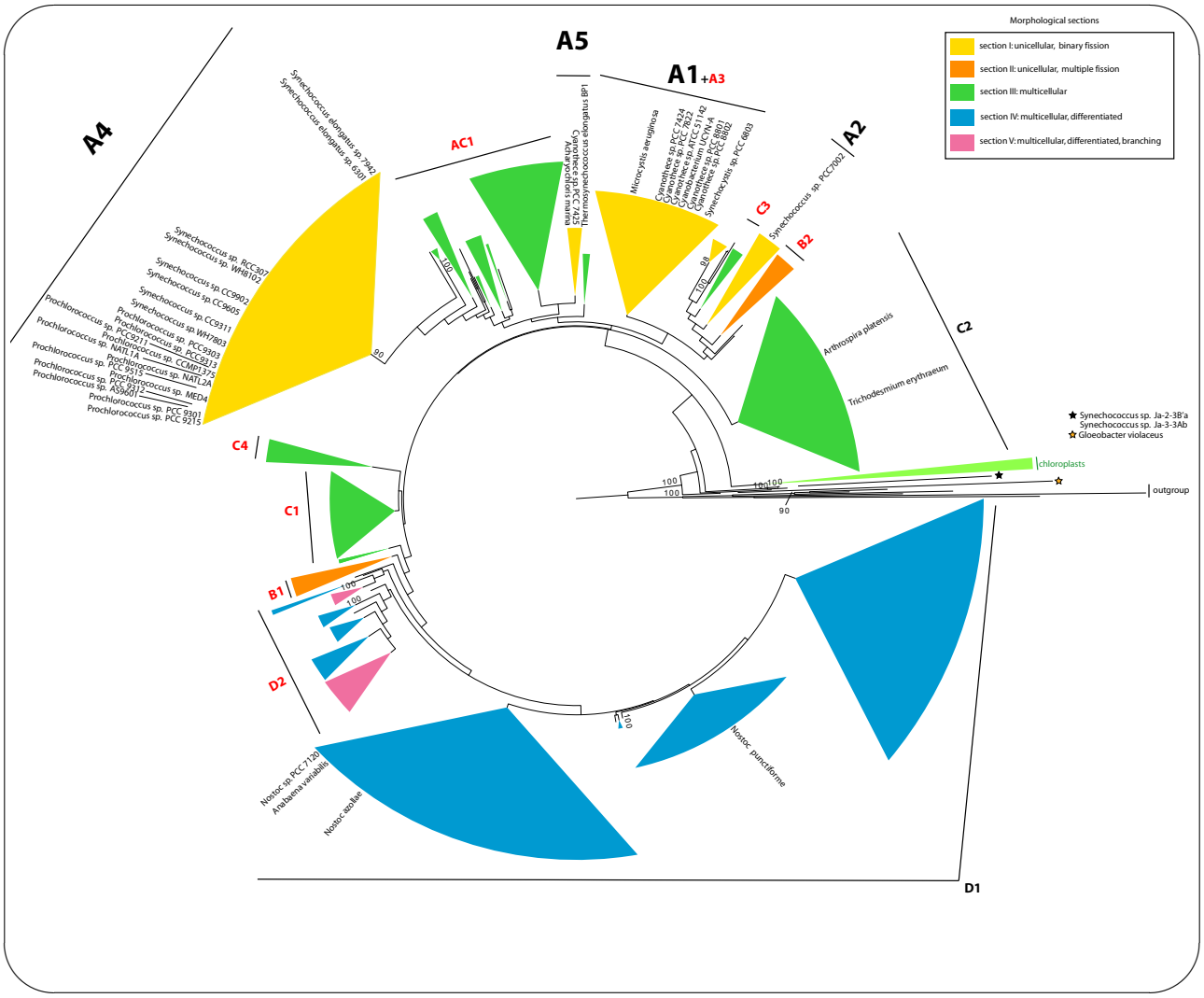


Fig. 3.1. Phylogenetic tree of 1,261 cyanobacterial species. Tree was reconstructed with maximum likelihood from 16S rRNA sequences. Morphological sections are marked by color. Fully sequenced strains are indicated on the tree by their taxonomic names. Phylogenetic groups (A1-A5;B1,B2;C1-C4;D1,D2) have been described elsewhere. Four groups for which no genome data are available at present are marked red. Taxa have been selected as previously described,4 with the addition of 41 taxa for which full genome sequences are available.

size of groups containing cultivated cyanobacteria (in terms of the number of taxa) does not necessarily represent its actual size in nature. In the reconstructed trees the taxonomic sampling from smaller groups is likely to contribute demonstrably to increased phylogenetic diversity (PD). Unfortunately no sequence data are available for those groups (AC1, B1, B2, C1, C3, C4 and D2). Morphological sections II and V lack genome data completely. Figure 3.2 shows a well resolved phylogenetic reconstruction of cyanobacterial taxonomy, highlighting in red the species recommended for future genome projects. Multicellular species are found in groups D1, D2, C1, C2, C3, C4 and AC1, but only six species from two of these groups (D1,C2) are completely sequenced. Special attention should be brought to multicellular species from groups AC1 and C4 which branch closer to the outgroup (Fig. 2) and therefore would be valuable for exploring the origin of multicellularity. Furthermore, genomes for future sequencing can be rigorously selected in order to maximize both the phylogenetic and morphological diversity. This can for example be done using a simple "greedy" algorithm [276],

by iteratively choosing the best immediate taxon (i.e., those adding most PD) based on the phylogeny depicted here.

State of the art methods for supermatrix and supertree construction aim to infer the species phylogeny from

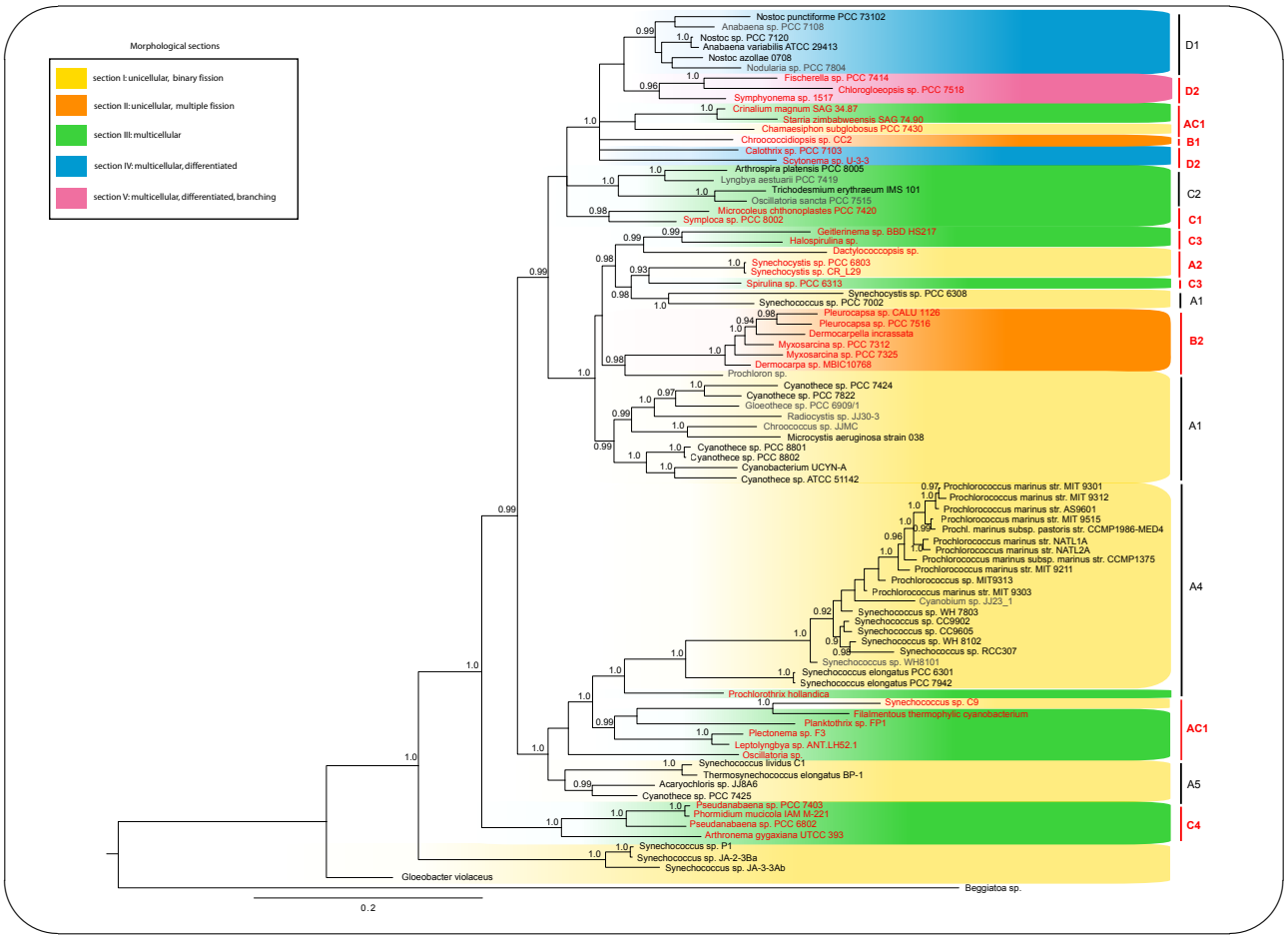
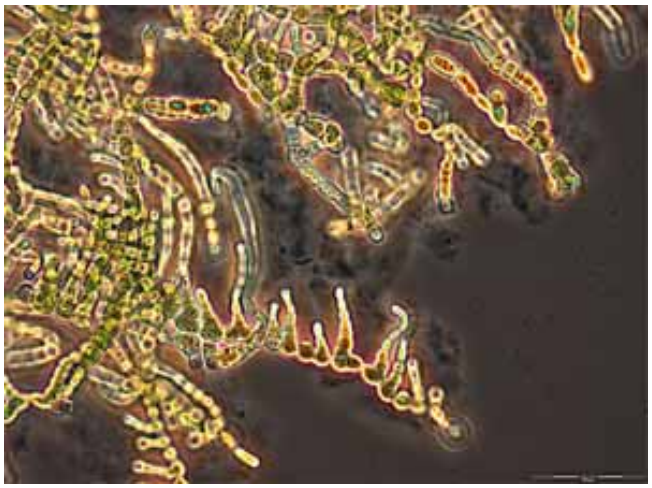


Fig. 3.2. Phylogenetic tree indicating where full genome data are lacking. Tree of a cyanbacterial subset representing the full morphological and genomic diversity. It was reconstructed from 16S rRNA sequences using a Bayesian approach. Colors describe different morphological sections. Fully sequenced taxa are shown in black. Shown in red are taxa of groups where no genome sequence data are available at present, and which we recommended for genome sequencing. Taxa have been selected as previously described [159], with the addition of 41 taxa for which full genome sequences are available.

thousands of taxa and multiple genomic loci [277]. Although for prokaryotes inferring a single species phylogeny may be particularly hampered by frequent lateral gene transfer, recombination and species delimitation problems [278]. Nonetheless the availability of better genomics data will make it possible to study the distribution of individual gene histories, from which the evolutionary mechanisms and their patterns may be inferred. Developments in statistical and computational methodology offer improved alignment strategies [**?**] and enable inference in the presence of recombination, incomplete lineage sorting and lateral gene transfer. In addition, the natural lineage heterogeneity may now be studied at the genomic scale using non-stationary or non-reversible Markovian models [279], leading to the better understanding of the biological forces driving lineage specific biases such as those affecting the evolutionary rates and molecular composition in diverse groups.

A more representative sampling of cyanobacterial species would provide a basis for future phylogenomic analyses that would offer insights into the early history of cyanobacteria. This would not only strengthen and further test the results of the study discussed here, but also enable others to identify and reconstruct the evolution of a wide set of genes that have been responsible for morphological and biochemical adaptations in this phylum.

CHAPTER IV



Chapter IV: Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event

Authors:

Bettina E. SCHIRRMEISTER, Jurrian DEVOS, Alexandre ANTONELLI and Homayoun C. BAGHERI

(submitted)

Cyanobacteria are one of the oldest and most diverse prokaryotic phyla, with morphotypes ranging from unicellular to multicellular forms with terminal differentiation. It has been suggested that cyanobacteria raised oxygen levels in the atmosphere around 2.45-2.32 billion years ago during the "Great Oxidation Event" (GOE), hence dramatically changing life on the planet. Yet, little is known about the temporal evolution of cyanobacterial lineages, and possible interplays between the origin of multicellularity, diversification and the rise of atmospheric oxygen. We estimated divergence times of extant cyanobacterial lineages under a Bayesian relaxed clock for a dataset of 16S rRNA sequences representing the entire diversity of this phylum. We tested whether the evolution of multicellularity overlaps with the GOE, and whether multicellularity is associated with significant shifts in diversification rates in cyanobacteria. Our results indicate an origin of cyanobacteria before the rise of atmospheric oxygen. The evolution of multicellular forms coincides with the onset of the GOE and an increase in diversification rates. These results suggest that multicellularity could have played a key role in triggering cyanobacterial evolution around the GOE.

Great Oxidation Event | Multicellularity | Early life | Origin of Cyanobacteria | Diversification rate

Introduction

Cyanobacteria, formerly known as blue-green algae, are one of the morphologically most diverse groups of prokaryotic organisms.

According to the fossil record, various distinct morphotypes attributed to cyanobacteria were already present over two billion years ago (Bya) [123, 103]. The phylum is thought to have existed as early as 2.45-2.32 Bya, based on the assumption that cyanobacteria were responsible for the rapid accumulation of atmospheric oxygen levels, referred to as the "Great Oxidation Event"(GOE) [181, 110, 182, 183, 111]. Despite the generally accepted time frame for the rise of cyanobacteria, surprisingly little is known about when morphological innovations such as multicellularity first appeared. It is also unclear what influence, if any, these innovations may have had on the diversification of the phylum. The assumed link between the rise of atmospheric oxygen and cyanobacteria is also poorly understood: did the GOE follow the first appearance of cyanobacteria, or did it follow at a later point, in possible association with morphological innovations of the phylum?

There have been previous attempts to estimate the origin of cyanobacteria and their morphotypes [143, 213, 120, 121]. However, it is likely that a biased taxonomic choice, especially towards the early branches of cyanobacterial phylogeny, may have led to incomplete conclusions [170, 55]. Phylogenetic evidence indicates that multicellularity evolved very early in the history of cyanobacteria, challenging the view that multicellularity is a derived condition in the phylum [159]. Nonetheless, important questions remain: (i) when did cyanobacteria and their major clades evolve? (ii) when did multicellularity first appear? (iii) how are these transitions associated with the GOE around 2.45-2.32 Bya?

The far-reaching impact of the GOE cannot be emphasized enough: it changed Earth's history by enabling the evolution of aerobic life. Unlike other eubacterial phyla, cyanobacteria exhibit a well studied fossil record [103, 280, 123, 107, 281]. However, fossil data are often limited and present only minimum age estimates of clades. Therefore, a combination of fossil data with molecular phylogenetic methods has been advocated [282, 283, 284]. Thus, the use of carefully selected calibration priors for molecular dating analyses can provide new insights into the temporal evolution of cyanobacteria and the early history of life. We estimated divergence times of cyanobacteria, while addressing different interpretations of the fossil record as calibration priors. We then evaluated whether the GOE coincided with the development of major cyanobacterial morphotypes present today. Furthermore, we tested for significant shifts in the rate of cyanobacterial diversification, incorporating information on the phylogenetic position of 281 species and 4,194 strains. Our results support theories of an early cyanobacterial origin towards the end of the Archean Eon, before the rise of oxygen. Evolution of multicellulartiy coincided with the GOE, and corresponded to a marked increase of diversification in cyanobacteria.

Results

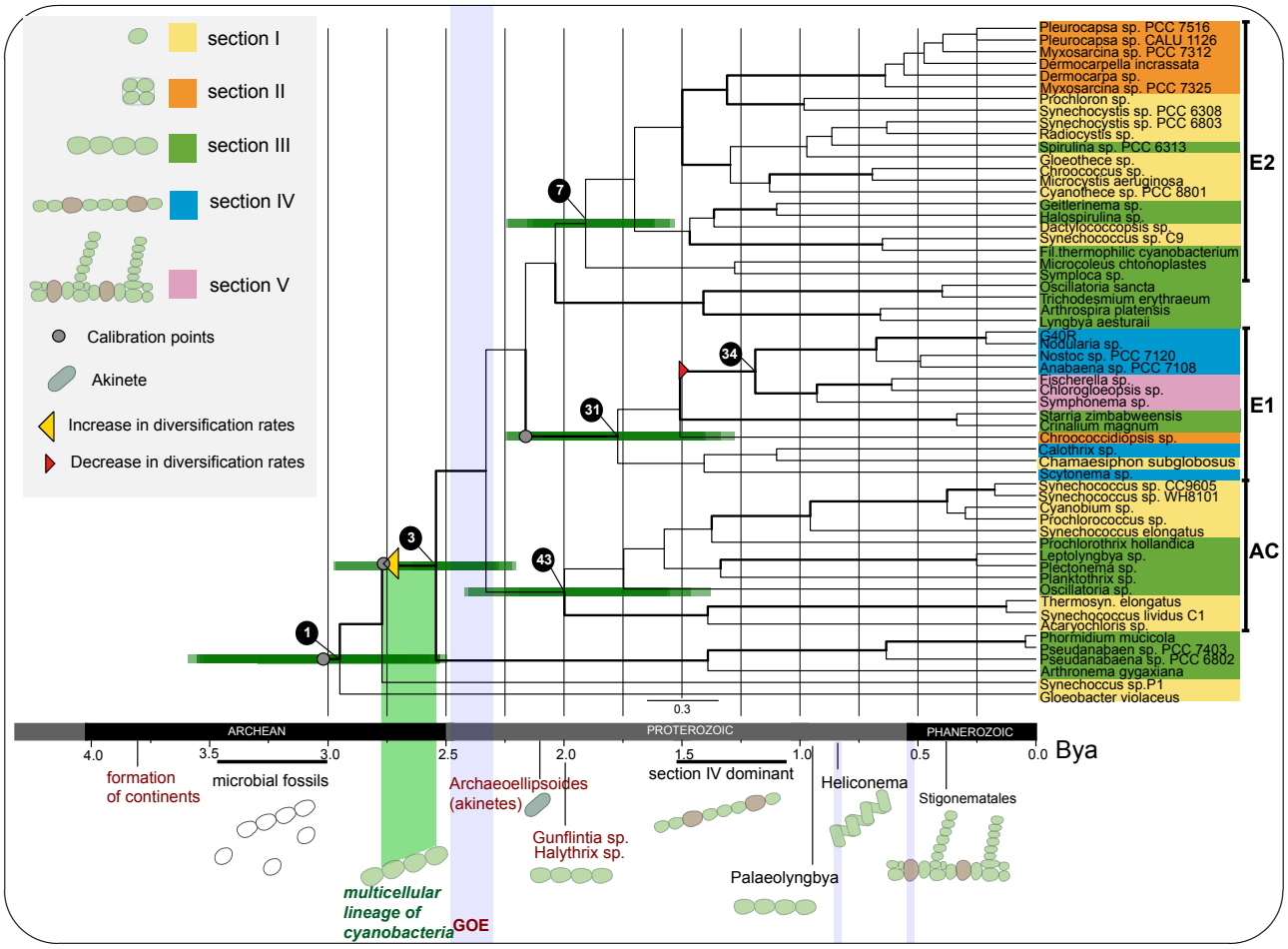


Fig. 4.1. Bayesian consensus tree of cyanobacteria displaying divergence time estimates. Phylogenetic tree with 95% confidence intervals of the discussed node ages shown as green bars. Morphological features of taxa are marked by colored boxes and listed in the inset. Branches with posterior probabilities > 0.9 in all analyses are presented as thick lines. Branches used for calibration of the tree are marked by grey circles. A significant increase in diversification rate (yellow triangle) —9.66 fold (average of all analyses)— can be detected at node 3 and a minor decrease (red triangle) at 33/34. The earlier shift close to node 3 coincides with the origin of multicellularity. Schematic drawings of cyanobacterial fossils are provided under the timeline, with the ones used for calibration of the tree marked in red. Our results indicate that multicellularity originated before or at the beginning of the GOE (green colored).

Phylogenetic analyses.

To infer the early evolution of cyanobacteria, we reconstructed Bayesian phylogenetic trees using 16S rRNA sequence data. Although genomic data are increasing, available cyanobacterial genome sequences are heavily biased towards unicellular species and therefore present neither the morphological nor the genetic diversity of this phylum [170]. In a previous study [159], a phylogenetic tree of 1,200 cyanobacterial sequences was reconstructed, from which a subset of taxa was sampled that represents the surveyed diversity of this phylum. Here we used this subset plus one strain (G40R) that represents a potentially new distinct species isolated by our group. The Bayesian consensus tree of analysis 7 is presented in Figure 4.1, including age estimates of important nodes as given by analyses 1, 3, 5, and 7 (Table 4.1). Our phylogenetic results agree with recent studies [112, 85, 115, 120, 159] and show that none of the morphological groups (section I-V) previously de-

scribed [64, 84] are monophyletic (Additional File 4.1; Fig. 4.1). Furthermore, *Gloeobacter violaceus* is resolved as the sister group of all other cyanobacteria. Three major clades can be distinguished (Clades E1, E2 and AC; Additional File 4.1, Additional File 4.2, Fig. 4.1) together representing the majority of cyanobacterial taxa living today (84% of 4,194 strains accounted for in this study). All clades have been defined in a previous publication [159], with clades E1 and E2 being subclades of a clade E that includes species of all morphological sections. Species belonging to morphological sections IV and V occur solely in E1. Clade AC contains marine pico-phytoplankton genera *Synechococcus* and *Prochlorococcus* (section I) as well as some undifferentiated multicellular species (section III).

Divergence time estimation

Divergence times along the cyanobacterial phylogeny were estimated under a Bayesian relaxed molecular clock [285]. Eight different analyses were performed to take a broad range of prior assumptions into account and evaluate their influence on the results (Table 4.1, Fig. 4.2). Median node ages ($\tilde{m}$) are shown in Figure 4.2, and together with 95% highest probability densities (HPD) are provided in Table 4.1 (discussed nodes) and Additional File 4.5 (all nodes). Although ages of cyanobacterial nodes varied with respect to the analyses, our major conclusions are robust to different calibration priors. All analyses indicated that cyanobacteria originated prior to the GOE (2.45 Bya). Multicellularity most likely originated along the branch leading to node 3 [159]. For this node, most of the analyses suggested a median age either before the GOE, or at the beginning of it. The ancestor of the lineage leading to node 3 was also a calibration point in our analyses (Table 4.1). Figure 4.3 compares the implied prior probability distributions of that calibration point to posterior probabilities of the node 3, in order to assess to what extent our prior assumptions affected the outcomes of the analyses. Although the prior assumptions put a higher probability on an age after the GOE around 2.2 Bya, our results suggest an older median node age for node 3, between 2.42-3.08 Bya (all analyses;Table 4.1), which is well before the GOE. Furthermore, the clades E1, E2 and AC are estimated to have originated around the end of the GOE. These clades comprise the majority of living cyanobacteria (77% of 281 species and 84% of 4,194 strains).

Shifts in Diversification Rates

In order to identify whether the GOE or multicellularity might have influenced the net diversification of cyanobacteria (i.e. speciation rate), we tested whether diversification rates have been constant through time, based on data for 281 species covering 4,194 strains. Since previous work suggested that taxonomy of cyanobacteria needed revision [64], we ran analyses incorporating information on both species and strains. Clades containing many species also contain many strains (Additional File 4.6). Therefore, results from the diversification rate estimation showed similar patterns independent of whether species numbers or strain numbers were used (Additional File 4.7). Two significant shifts of diversification rates were detected: At node 3 , where multicellularity evolved, the diversification rate increased 9.66 fold (averaged over all analyses; Additional File 4.7). Subsequently at node 33/34, the diversification decreased slightly. Both results were recovered in 15 out of 16

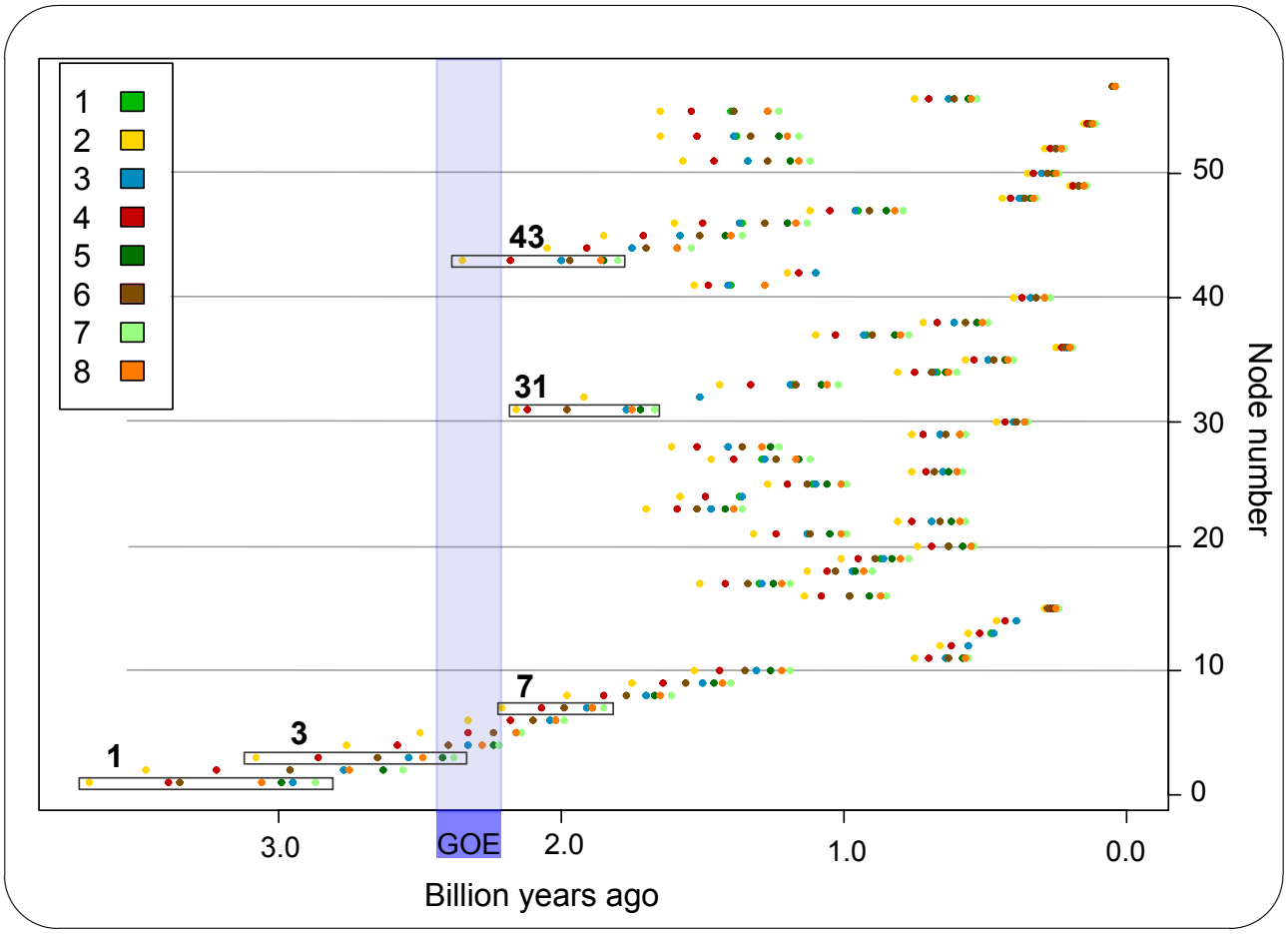


Fig. 4.2. Median age estimates under eight analytical scenarios. Median age estimates of clades (Table 4.1). The origin of cyanobacteria (node 1) and the evolution of multicellularity (node 3) are estimated before or at the beginning of the Great Oxidation Event (GOE). Relatively soon after the GOE, the stem lineages of the three major cyanobacterial clades originated, containing unicellular cyanobacteria (node 7), terminally differentiated taxa (node 31), and marine phycoplankton (node 43).

analyses (Additional File 4.3 and 4.4; Fig. 4.1).

Discussion

Evolution of multicellularity and possible consequences

Our results indicate a concurrence of the origin of multicellularity, rise of oxygen and an increased diversification rate of cyanobacteria. Posterior probability densities of our divergence time analyses point to a potential origin of multicellular cyanobacteria before or at the beginning of the GOE. The transition to multicellularity represents an important change in organismic complexity [**?**]. There are various advantages that multicellularity could confer [185, 286]. Among others, increased size protects from predation, and cooperation of cells may also increase fitness due to economies of scale. Growth in a multicellular setting presupposes intra- and inter-cellular communication, which has been found in cyanobacteria [189, 191]. It also requires cellular

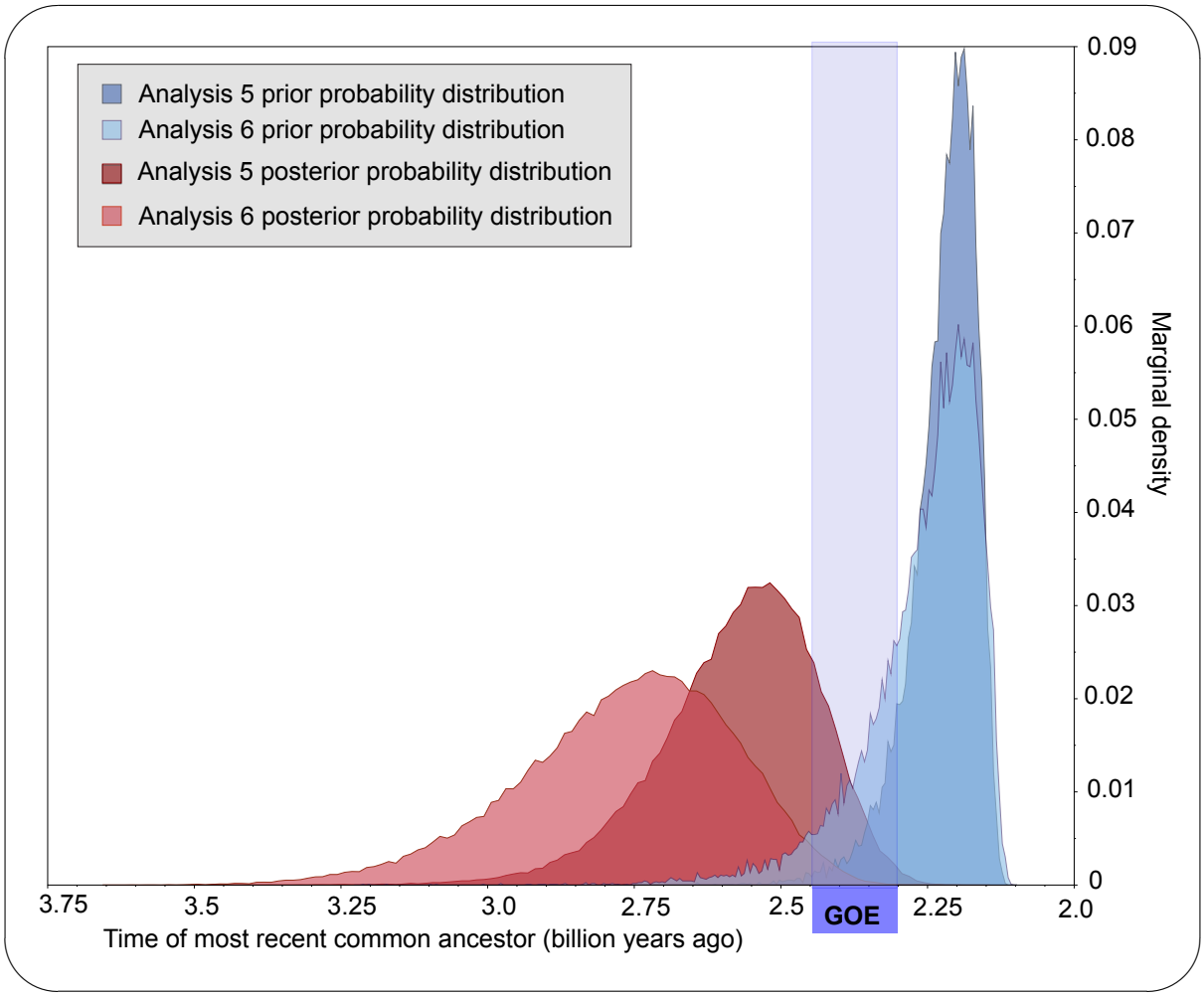


Fig. 4.3. Prior and posterior probability densities of Node 3 Marginal prior probability densities of analyses employing narrow (5) and wide (6) prior distributions were conservatively biased towards younger ages, strongly favoring an origin of multicellularity after the GOE. However, posterior probabilities point to an origin of multicellularity before or at the beginning of the GOE. Marginal prior probability distributions were estimated in analyses that only sampled from the prior.

recognition of polarity [224]. Branching growth, as can be observed in cyanobacteria belonging to section V, potentially requires even more complex cell communication to coordinate cell growth.

Recent studies have shown that multicellularity might evolve relatively fast given selective pressure [287] and may provide metabolic fitness advantages compared to single cells [288]. Increased fitness can promote an abundance of multicellular cyanobacteria, spurring higher oxygen production. The accumulation of atmospheric oxygen may have led to formation of new ecological opportunities. Increased diversification rates around the time when multicellularity evolved point to the possibility that cyanobacteria might have used new adaptive opportunities. At the end of the GOE, three clades (E1, E2 and AC) evolved that represent the majority of cyanobacteria living today.

Analysis	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Model assumptions and calibration points								
Outgroup	-	-	yes	yes	yes	yes	-	-
Root	-	-	exp	exp	exp (2.45;2.816)*	exp (2.45;2.816)*	(2.45-3.8) Bya	(2.45-3.8) Bya
Node 3	LN $(2.1; 2.27, 0.5)$	LN $(2.1; 2.58, 0.8)$	LN $(2.1; 2.27, 0.5)$	LN $(2.1; 2.58, 0.8)$	LN $(2.1; 2.27, 0.5)$	LN $(2.1; 2.58, 0.8)$	LN $(2.1; 2.58, 0.8)$	LN $(2.1; 2.58, 0.8)$
Node 31 or 32	exp (2.1;2.13,1)	exp (2.1;2.13,1)	exp (2.1;2.13,1)	exp (2.1;2.13,1)	exp (2.1;2.13,1)	exp (2.1;2.13,1)	exp (2.1;2.13,1)	exp (2.1;2.13,1)
Results for discussed nodes								
	($\tilde{m}$) (HPD)	($\tilde{m}$) (HPD)	($\tilde{m}$) (HPD)	($\tilde{m}$) (HPD)	($\tilde{m}$) (HPD)	($\tilde{m}$) (HPD)	($\tilde{m}$) (HPD)	($\tilde{m}$) (HPD)
Node 1	2.95 (2.5-3.6)	3.67 (2.79-4.74)	2.99 (2.57-3.55)	3.35 (2.74-4.15)	2.87 (2.53-3.30)	3.06 (2.66-3.53)	2.95 (2.53-3.55)	3.39 (2.87-3.80)
Node 3	2.54 (2.28-2.98)	3.08 (2.42-3.84)	2.42 (2.21-2.73)	2.65 (2.28-3.18)	2.38 (2.20-2.62)	2.49 (2.26-2.81)	2.54 (2.29-2.97)	2.86 (2.43-3.34)
Node 7	1.91 (1.62-2.25)	2.21 (1.74-2.78)	1.89 (1.57-2.17)	1.99 (1.63-2.41)	1.85 (1.53-2.13)	1.89 (1.56-2.21)	1.91 (1.62-2.24)	2.07 (1.71-2.50)
Node 31	1.77 (1.4-2.24)	2.16 (1.53-2.56)	1.72 (1.34-2.20)	1.98 (1.39-2.34)	1.67 (1.28-2.17)	1.75 (1.30-2.23)	1.77 (1.41-2.25)	2.12 (1.50-2.41)
Node 43	2 (1.56-2.43)	2.35 (1.73-3.03)	1.85 (1.46-2.25)	1.97 (1.48-2.50)	1.80 (1.38-2.19)	1.86 (1.41-2.30)	2.00 (1.57-2.41)	2.18 (1.71-2.72)

Table 4.1: Eight different combinations of calibration priors for the divergence time estimation. Exp: exponential distribution (offset;mean). $LN()$: lognormal distribution (offset;mean,stdev). * truncated at 3.8 Bya.

Early Earth history and the fossil record

Our finding that cyanobacteria may have been around longer than previously anticipated is congruent with reconstructions of the early history of Earth. The origin of Earth is deduced to date back ∼4.5 Bya [289, 252]. Subsequently, the planet cooled down and eventually separated into core, mantle, and crust [290]. The oldest indications for continental crust come from 4.4 B. yrs. old zircon crystals [253]. Further indications for continental crust come from granitic remnants estimated to be 4.0 B. yrs. old [252]. From 3.8 Bya onwards, continents are assumed to have built up gradually over time, possibly exhibiting high growth rates [252]. Permanent existence of life prior to 4.2-3.8 Bya is unlikely considering that the young Earth was subject to strong bombardment by asteroids [291, 289]. Earliest physically possible existence of durable lifeforms is dated around 3.8 Bya [289, 252], although fossil evidence does not predate ∼3.45 B. yrs. [196, 197]. Most of these prokaryotic fossils from the early Archean Eon have been identified in two regions: the Berberton Greenstone Belt (BGB), South Africa (around 3.20-3.50 B. yrs. old) and the Pilbara Craton (PC), Western Australia (around 2.90-3.60 B. yrs. old) [195, 243, 196, 244, 245, 246, 197]. The oldest fossils from these regions, around 3.45 B. yrs. old, are spherical, probably hyperthermophilic microbes (Hooggenoeg Formation, BGB [245, 197]) and filaments of possibly anoxygenic photosynthetic prokaryotes (Panorama Formation, East-PC [196, 197]), both around 3.45 B. yrs. old. Further evidence for life includes 3.4 B. yrs. old trace fossils (PC) [246], 3.416 B. yrs. old deformed microbial mats (BGB), sedimentary structures generated from microbes (BGB) [243], or 3.0 B. yrs. old biofilms (PC) [244].

The first indisputable cyanobacterial fossils date back around 2.0 Bya and come from two localities, the Gunflint iron formation and the Belcher Subgroup (both in Canada) [107, 281]. Although differences in the microbial fossil composition have been recognized [107], both cherts include filamentous and coccoidal species. *Gunflintia grandis* and *G. minuta* have been identified as filamentous cyanobacterial fossils from the Gunflinta iron formation, and *Halythrix sp.* has been described as an oscillatorian fossil from the Belcher subgroup [103] (Fig. 4.1). Cyanobacterial fossils younger than 2.0 B. yrs. are more widely distributed. Various examples are given in Figure 4.1. Terminally differentiated cyanobacteria (section IV and V) are dominant during the Mesoproterozoic [281]. Results from our molecular dating analyses shift the origin of unicellular, as well as multicellular cyanobacteria to before the "Great Oxidation Event". Archean fossil findings may potentially depict remains of cyanobacteria, but because they cannot be assigned beyond doubt, they have been given the status of "first possible" and "first probable" remnants [281]. "Possible" cyanobacterial fossils have been found in 2.52-2.55 B. yrs. old cherts of the Transvaal Supergroup in South Africa [281, 292]. "Probable" unicellular and filamentous cyanobacterial fossils are distributed in cherts from the Archean Eon, around 2.6 Bya [281, 248, 293, 294] and 3.26 Bya [294]. Although previously described biomarkers that supported an existence of cyanobacteria around 2.7 Bya [295, 296] have been dismissed [184], new evidence has been found in favor of an early cyanobacterial origin [297, 298, 299]. Our molecular dating results place the origin of both unicellular and multicellular cyanobacteria to prior to the GOE, and thus suggest that some of those fossils could indeed represent relatives of cyanobacterial lineages.

Recent studies have suggested that oxygen accumulation in oceans, occurred approximately 200-300 million years before the GOE [298, 297]. Current evidence from the fossil record, geochemical findings, and our molecular analyses, together suggest a distinct origin of cyanobacteria before the GOE. The origin of multicellularity towards the GOE could have had positive fitness effects leading to an increase in cyanobacterial abundance and subsequently positively influencing net oxygen production.

Conclusion

Cyanobacteria are one of the oldest, morphologically most diverse prokaryotic phyla on this planet. It is widely accepted that they caused the " Great Oxidation Event"(GOE) starting 2.45 Bya, but debates about their origin are still ongoing [105, 200, 184]. Various lines of fossil and geochemical evidence have accumulated supporting an origin of cyanobacteria before 2.45 Bya. [281, 248, 294, 299, 297, 298]. Here, we used Bayesian phylogenetic analyses using relaxed molecular clocks and different combinations of calibration priors. We estimated the origin of cyanobacteria and their dominant morphotypes with respect to the GOE. Though resulting age estimates of the different analyses differ somewhat in their HPD, robust statements regarding the origin of cyanobacteria and their morphotypes can nevertheless be formulated: (i) cyanobacteria seem to have originated before the GOE, (ii) multicellularity coincides with the beginning of the rise of oxygen, and (iii) three clades representing the majority of extant cyanobacteria evolved shortly after the accumulation of atmospheric oxygen.

Methods

Taxon sampling

The majority of sequences was downloaded from GenBank [300] (Additional File 4.6). Three eubacterial species were chosen as an outgroup, *Beggiatoa sp.*, *Chlamydia trachomatis* and *Spirochaeta*. *Beggiatoa sp.* shows the closest 16S rRNA sequence distance to cyanobacteria [159], although full multilocus analyses have suggested *Deinecococcus* to be closest relative of cyanobacteria [212].

A total of 58 cyanobacterial species were chosen for the analyses. Aside from strain "G40R" (Additional File 4.8) all taxa were selected as described in a previous study [159]. The taxa chosen comprise representatives from all morphological sections described by Castenholz *et al.* [64] and represent the full morphological and genetic diversity of this phylum [159]. Strain "G40R" (Additional File 4.8) is a yet uncharacterized, terminally differentiated, multicellular isolate from the North Sea. It was isolated from ponds at the shore of North-Western Ameland, Netherlands. The strain was then cultivated in ASN III seawater medium and kept at $15\,^{\circ}C$ in an environmental chamber at a constant day/night cycle of 6 hours darkness and 18 hours light.

Alignment and Phylogenetic analyses

All phylogenetic analyses were based on 16S rRNA gene sequences selected from a variety of cyanobacterial strains. For most of those, full genome data are not yet available. Multiple sequence alignments (msa) were constructed using the programs Clustal-X [175] and MUSCLE [301]. The estimated msa were compared visually with 'MSA Comparator' from the SuiteMSA package [302]. Alignments showed strong similarities. Further phylogenetic analyses were conducted on the MUSCLE alignment, after excluding sites with gaps. Analyses were performed on datasets with outgroups (i: 61 taxa, 1090 nucleotides) and without outgroups (ii: 58 taxa, 1077 nucleotides). Uncorrected and corrected Akaike Information Criterion (AIC)[303, 304], implemented in jModelTest v0.1.1 [305], indicated that the general time reversible substitution model with gamma distributed rate variation among sites (GTR+G) [267] was the most suitable model of sequence evolution. Phylogenetic relationships were estimated using MrBayes v.3.1.2 [176]. We employed two Markov Chain Monte Carlo runs, each calculating 6 Metropolis-coupled chains for 100 million generations sampling every 2,000th generation. Default priors were adequate and left unchanged, but the temperature parameter was adjusted to 0.1 to ensure proper mixing. Convergence between runs was achieved, as the potential scale reduction factor had approached 1.00 and average standard deviations of split frequencies was < 0.01. Mixing and convergence of all parameters was further assessed using the software Tracer v1.5 [177]. We combined runs after discarding the first 25% of samples as a conservative burnin, including only samples from the stationary phase. Effective sample sizes were large, $> 3,000$ for the likelihood samples and all estimated parameters, supporting a well mixed analysis.

Divergence time estimation

Preliminary analyses indicated strong non-clocklikeness of our data. Therefore, we applied an uncorrelated lognormal relaxed clock [306]. The analyses were conducted with and without outgroup, and a combination of three calibration points (Table 4.1). Additionally, monophyly constrains were set for three nodes that were well-supported in our MrBayes analyses (Supporting information): (i) the phylum cyanobacteria, (ii) cyanobacteria, excluding *Gloeobacter*, and (iii) cyanobaceria excluding *Synechococcus* P1 and *Gloeobacter* (Fig. 1). (i) has been extensively investigated and confirmed before, i.e. that cyanobacteria is a monophyletic group within the eubacteria [159]. (ii) An early divergence of Gloeobacter has been supported in previous analyses [96, 85, 170, 159]. Unlike other cyanobacteria, *Gloebacter violaceus* lacks thylacoid membranes [113], and various differences in gene content compared to cyanobacteria have been found [160]. (iii) *Synechococcus* P1 is a thermophilic unicellular cyanobacterium isolated from Octopus Spring in Yellowstone nationalpark [307]. The proximity to *Gloebacter* and eubacterial outgroups has been shown by genetic comparisons and phylogenetic analyses in previous analyses [85, 170, 159]. Divergence time estimation was conducted using the software BEAST v1.6.2 [306] and run on the CIPRES Science Gateway v.3.1 [264]. For each analysis, we ran 6 MCMC chains for 50 million generations, sampling every 2,000th generation. Although convergence across runs of all parameters was reached before 5 million generations, we used a conservative 25% burnin. The maximum clade credibility tree was calculated using TreeAnnotator [306], after combining runs. Results are presented on the 50% consensus tree calculated with SumTrees Version 3.3.1 by Jeet Sukumaran and Mark T. Holder, a part of the DendroPy library [308].

Calibration points

The root - Stem lineage of cyanobacteria

Four of the eight divergence time analyses included three eubacterial taxa as an outgroup (Table 4.1: 3,4,5,6). This enabled calibrating the cyanobacterial stem lineage. The GOE dates back 2.32-2.45 B. yrs. [110], and is assumed to be a result of cyanobacterial activity. For the analyses we therefore assumed that cyanobacteria must have existed when the GOE occurred. We use the start of the GOE as the latest date for the divergence of cyanobacteria and the outgroup [309]. The possibility of permanently existing lifeforms is suggested to occur earliest around 3.8 Bya [289, 252] which we use as earliest date of our root calibration. See Table 4.1 for detailed description of prior age probability distributions.

Node 3 - First multicellular cyanobacteria

Node 3 in Figure 4.1 has been estimated to be a multicellular ancestor of extant cyanobacteria, as recovered in a previous study [159]. Fossil records indicate that terminally differentiated cyanobacteria belonging to section IV and V evolved before 2.1 Bya. Such differentiation may only evolve in a multicellular setting [255]. We therefore assume that the stem lineage of node 3 must have been present prior to 2.1 B. yrs., and use this as a hard minimum bound of a lognormal prior distribution. We used a soft upper bound, linking the distribution

of prior probability mass to the timing of the GOE. Multicellularity may have evolved as a consequence of new habitats that became available after the GOE, 2.3 Bya, or it could instead have triggered a rise of oxygen in the atmosphere. Therefore, we distinguish two calibration scenarios, one by setting the probability of the age of node 3 to a lognormal distribution with 95% being younger than 2.45 (Table 4.1:1,3,5;narrow), and the other by setting the median age of the prior at 2.45 Bya (Table 4.1:2,4,6;wide).

Node 31 or 32 - First terminally differentiated cyanobacteria

Cyanobacteria belonging to section IV and V share the property to form resting cells named akinetes. These persistent cells are generated if environmental conditions are unfavorable. Fossilized remains of these akinetes have been identified at various locations throughout the Proterozoic [310, 311, 122, 123, 107]. The oldest of these fossilized akinetes are found in 2.1 B. yrs. old rocks [123, 143], and imply that terminally differentiated cyanobacteria belonging to section IV and V originated prior to 2.1 Bya. Furthermore, taxa of this group are capable of terminal differentiation. Oxygen sensitive nitrogen fixation is spatially separated from oxygenic photosynthesis and takes place in so called heterocysts. Oxygen levels providing a selective advantage for separation of these processes were reached approximately 2.45 Bya [143]. As a calibration for the divergence time estimation, we set the most recent common ancestor of sampled taxa of section IV and V at 2.1 B. yrs. as a hard minimum bound, and specified 95% of prior probability mass before 2.45 Bya, using an exponential distribution.

Shifts in Diversification Rates

To test whether the rate of lineage accumulation has been constant throughout cyanobacterial evolution, we used the function MEDUSA from the geiger 1.3-1 package in R [312]. MEDUSA uses maximum likelihood to estimate a birth-death model of diversification that includes the optimal number of rate shifts, while penalizing for excess parameters based on AIC-scores. We corrected for possible taxon sampling biases by including information on known numbers of extant species and strains, which were collected from GenBank (Additional File 4.6). This is a useful strategy because descriptions of cynobacterial species and strains typically include a 16S rRNA gene sequence, and strain and species numbers greatly differ among clades. Phylogenetic positions of unsampled species and strains in the cyanobacterial phylum were estimated with help of a phylogenetic tree of 1,200 taxa compiled in a previous study[159]. Subsequently, numbers of unsampled species and strains were assigned to taxa sampled for the dating analyses of this study (Additional File 4.6). MEDUSA was run based on 50% consensus trees calculated with SumTrees Version 3.3.1 [308], of the eight phylogenetic analyses from the divergence time estimation (Table 4.1). Inference based on maximum clade credibility trees gave qualitatively similar results.

CONCLUSION

The evolution of morphological complexity in cyanobacteria

Cyanobacteria are one of the oldest and most diverse prokaryotic organism on the planet. They are assumed to be responsible for the "Great Oxidation Event" (GOE), a key event for the evolution of aerobic life on Earth [110, 182]. Morphologically they exhibit impressive features (unique among prokaryotes) which have caused a misclassification of this group as part of the plant kingdom [57, 84, 68]. It is known from a well preserved fossil record that many of these diverse features were already present in ancient taxa, 2.0 billion years ago [103, 123]. Yet, we do not have a resolution to the question of how morphological characteristics like multicellularity evolved, and what impact this might have had for the evolution of this phylum and life on the planet. Along other lines of evidence [58], molecular phylogenetic studies supported a classification of cyanobacteria among prokaryotes. Molecular phylogenetic tools, built on statistical frameworks like Parsimony, Maximum Likelihood and Bayesian inference to reveal that morphological features on their own are not sufficient to deduce the history of this phylum [112, 85]. Applying molecular phylogenetic methods, I was able to reconstruct parts of the early evolution of cyanobacterial morphotypes.

Comparison of full genome data to morphological features indicated a correlation of increased ribosomal operon copy numbers with the occurrence of terminal cell differentiation. Furthermore, multiple, very strongly conserved 16S rRNA gene copies seem to be a result of a combination of concerted evolution driven by homologous recombination, and purifying selection. The close relation of photosynthetically-"primitive" *Gloeobacter violaceus* and Synechococci species isolated from hot springs was confirmed by a shared pattern of multiple

gene copy numbers. Variation of 16S rRNA sequences within cyanobacteria is noticeably low compared to other eubacterial phyla. These results support the potential of 16S rRNA genes as a taxonomic marker for phylogenetic analysis at a genus level in cyanobacteria.

I have shown that genome data recovered so far, represent neither the morphological nor genetic diversity of this phylum. To date, phylogenomic analyses miss important taxa for the inference of early evolutionary histories in cyanobacteria. Using a well chosen subset that covers the complete diversity of this phylum enabled the reconstruction of the origin and evolution of multicellularity in cyanobacteria. Living species of cyanobacteria descend from a unicellular most recent common ancestor. *Gloeobacter violaceus* and thermophilic Synechococci strains are the only extant lineages evolving directly from this ancestor. Multicellularity originated early during the history of cyanobacteria, and led to the majority of morphological variety, including four of the five previously defined subsections described by Rippka *et al.* [84, 68]. Subsequently, multicellularity was lost at least five times and regained at least once within this phylum.

The incorporation of fossil data enabled a combination of various calibration assumptions for divergence time estimation. This could confirm previous studies [297, 298] which point towards an origin of cyanobacteria distinctly before the GOE. Multicellularity in cyanobacteria originated shortly before or at the beginning of the GOE. Following the GOE, three clades evolved that together make up 70-80% of known living cyanobacetria. Furthermore, the evolution of multicellularity coincided with a distinct increase in diversification rates. It appears that the evolution of multicellularity in cyanobacteria might have conferred potential fitness advantages, resulting in increased abundance which could have had effects on oxygen production. The development of new habitats as a consequence of atmospheric oxygen increase could have provided opportunities for cyanobacteria.

Cyanobacteria have had a distinct impact on the history of Earth, providing the basis for the evolution of aerobic life. It seems that the rise of multicellularity was involved in this event, occurring more than 2.4 billion years ago. An attentive sampling of cyanobacterial species for genome analysis and the incorporation of data from the fields of palaeontology, geochemistry and physiology may offer the possibility for further analyses regarding the diversification of cyanobacterial morphotypes. In order to comprehend the importance of this phylum for the evolution of life, the exploration of blue-green algae needs to be continued.

Bibliography

[1] Bonnet C: *Oeuvres d'histoire naturelle et de philosophie.* Chez Samuel Fauche, Libraire du Roi, Neuchatel 1779. [Digitized by the Internet Archive in 2009 with funding from Lyrasis Members and Sloan Foundation].

[2] Kutschera U: **From the scala naturae to the symbiogenetic and dynamic tree of life**. *Biology Direct* 2011, **6**:33.

[3] Thompson DW: *Historiae Animalium - The Works of Aristotle*. Clarendon Press, Oxford 1910.

[4] Wehner R, Gehring W: *Zoologie*. Georg Thieme Verlag 1995.

[5] Ragan MA: **Trees and networks before and after Darwin**. *Biology Direct* 2009, **4**:43.

[6] Haeckel E: *Generelle Morphologie der Organismen*. Druck und Verlag von Georg Reimer 1866.

[7] Lamarck JB: *Philosophie zoologique ou exposition des considérations relatives à l'histoire naturelle des animaux*. Dentu Libraire 1809.

[8] Mayr E: *The Growth of Biological Thought*. The Belknap Press of Harvard University Press 1982.

[9] Wallace AR: **On the Law which has regulated the Introduction of new Species**. *Proceedings of the Linnean Society of London* 1960, **171(2)**:141–153.

[10] Lyell C: *Principles of Geology*. John Murray, Albermarle Street, London 1830.

[11] Wallace AR, Darwin CR: **On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection**. *Journal of the Proceedings of the Linnean Society of London. Zoology* 1858, **3(9)**:45–62.

[12] Darwin CR: *The Origin of Species by Means of Natural Selection*. John Murray, Albemarle Street, London 1859.

[13] Mayr E, Provine WB: *The Evolutionary Synthesis*. Harvard University Press 1980.

[14] Provine WB: *The Origins of Theoretical Population Genetics*. University of Chicago Press 2001.

[15] Ruse M: **Julian Huxley on Darwinian evolution: A snapshot of a theory**. *Metascience* 2011, **20**:329–333.

[16] Dayrat B: **The roots of phylogeny: How did Haeckel build his trees?** *Systematic Biology* 2003, **52**(4):515–527.

[17] Bowler PJ: *Evolution: History In:Encyclopedia of Life Sciences (ELS), Volume 38*. John Wiley & Sons, Ltd: Chichester 2010.

[18] Bowler PJ: *The Eclipse of Darwinism: Anti-Darwinian Evolution Theories in the Decades Around 1900*. The John Hopkins university Press 1992.

[19] Morgan TH: *A Critique of the Theory of Evolution*. Princton university Press 1916.

[20] Fisher RA: **The Correlation Between Relatives on the Supposition of Mendelian Inheritance**. *Philosophical Transactions of the Royal Society of Edinburgh* 1918, **52**:399–433.

[21] Fisher RA: *The Genetical Theory of Natural Selection*. At the Clarendon Press 1930.

[22] Wright S: **Evolution In Mendelian Populations**. *genetics* 1931, **16**:97–159.

[23] Haldane JBS: *The causes of Evolution*. Princeton Science Library 1990.

[24] Kutschera U, Niklas KJ: **The modern theory of biological evolution: an expanded synthesis**. *Naturwissenschaften* 2004, **91**:255–276.

[25] Wallace AR: **On the tendency of varieties to depart indefinitely from the original type**. *Journal of the Proceedings of the Linnean Society of London. Zoology* 1858, **3**:46–50.

[26] Weismann A: *Essays Upon Heredity*. Clarendon Press, Oxford 1889.

[27] Dobzhansky T, Sturtevant AH: **Inversions in the Chromosome of Drosophila pseudoobscura**. *Genetics* 1938, **23**:28–64.

[28] Florkin M: **The Call for Comparative Biochemistry**. *Comp. Biochem. Physiol.* 1970, **44B**:1–10.

[29] Hennig EHW: *Grundzuege einer Theorie der phylogenetischen Systematik,*. Berlin, Deutscher Zentralverlag 1950.

[30] Hennig EHW: *Phylogenetic Systematics*. University of Illinois Press 1966.

[31] Dupuis C: **Hennig,willi Impact On Taxonomic Thought**. *Annual Review of Ecology and Systematics* 1984, **15**:1–24.

[32] Sneath PHA, Sokal RR: *Principles of Numerical Taxonomy*. W. H. Freeman and Company 1963.

[33] Simpson GG: **Numerical Taxonomy and Biological Classification.** *Science (New York, N.Y.)* 1964, **144**(3619):712–3.

[34] Mayr E: **Numerical Phenetics and Taxonomic Theory**. *Systematic Zoology* 1965, **14(2)**:73–97.

[35] Zimmermann W: **Phanetische Und Phylogenetische Verwandtschaft**. *Phyton-annales Rei Botanicae* 1966, **11**(3-4):145–&.

[36] Hamilton A, Wheeler QD: **Taxonomy and Why History of Science Matters for Science**. *The University of Chicago Press and The History of Science Society* 2008, **99(2)**:331–340.

[37] Susko E, Inagaki Y, Roger AJ: **On inconsistency of the Neighbor-Joining, least squares, and minimum evolution estimation when substitution processes are incorrectly modeled**. *Molecular Biology and Evolution* 2004, **21**(9):1629–1642.

[38] Sanger F: **The Terminal Peptides of Insulin**. *Biochemical Journal* 1949, **45**(5):563–574.

[39] Harris JI, Sanger F, Naughton MA: **Species Differences In Insulin**. *Archives of Biochemistry and Biophysics* 1956, **65**:427–438.

[40] Zuckerkandl E, Pauling L: **Molecules as Documents of Evolutionary History**. *J. Theoret. Biol.* 1965, **8**:357–366.

[41] Mayr E: **Cause and Effect In Biology - Kinds of Causes, Predictability, and Teleology Are Viewed By A Practicing Biologist**. *Science* 1961, **134**(348):1501 – 1506.

[42] Dietrich MR: **The Origin of the Neutral Theory of Molecular Evolution**. *Journal of the History of Biology* 1994, **27(1)**:21–59.

[43] Fitch WM, Margoliash E: **Construction of Phylogenetic Trees**. *Science* 1967, **155**(3760):279 – 284.

[44] Fitch WM: **Distinguishing Homologous From Analogous Proteins**. *Systematic Zoology* 1970, **19**(2):99 – 113.

[45] Sanger F, Coulson AR: **Rapid Method For Determining Sequences In Dna By Primed Synthesis With Dna-polymerase**. *Journal of Molecular Biology* 1975, **94**(3):441 – 448.

[46] Sanger F, Nicklen S, Coulson AR: **Dna Sequencing With Chain-terminating Inhibitors**. *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**(12):5463–5467.

[47] Maxam AM, Gilbert W: **New Method For Sequencing Dna**. *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**(2):560–564.

[48] Woese CR, Fox GE: **Phylogenetic Structure of Prokaryotic Domain - Primary Kingdoms**. *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**(11):5088–5090.

[49] Woese CR: **Bacterial Evolution**. *Microbiological Reviews* 1987, **51**(2):221–271.

[50] Felsenstein J: **Maximum-likelihood Estimation of Evolutionary Trees From Continuous Characters**. *American Journal of Human Genetics* 1973, **25**(5):471–492.

[51] Felsenstein J: **Maximum Likelihood and Minimum-steps Methods For Estimating Evolutionary Trees From Data On Discrete Characters**. *Systematic Zoology* 1973, **22**(3):240–249.

[52] Felsenstein J: **Evolutionary Trees From Dna-sequences - A Maximum-likelihood Approach**. *Journal of Molecular Evolution* 1981, **17**(6):368–376.

[53] Yang ZH, Rannala B: **Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo method RID C-1358-2008**. *Molecular Biology and Evolution* 1997, **14**(7):717–724.

[54] Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R: **Microbiology in the post-genomic era RID G-4703-2011**. *Nature Reviews Microbiology* 2008, **6**(6):419–430.

[55] Wu DY, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng JF, Lucas S, Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk HP, Eisen JA: **A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea RID G-9608-2011**. *Nature* 2009, **462**(7276):1056–1060.

[56] Sapp J: **The prokaryote-eukaryote dichotomy: Meanings and mythology**. *Microbiology and Molecular Biology Reviews* 2005, **69**(2):292–+.

[57] Geitler L: *Kryptogamen-Flora - Die Algen - Cyanophyceae von Europa*. Akademische Verlagsgesellschaft m. b. H. 1932.

[58] Stanier RY: **The position of cyanobacteria in the world of phototrophs**. *Crlsberg Res. Commun.* 1977, **42**:77–98.

[59] Drews G: **The roots of microbiology and the in£uence of Ferdinand Cohn on microbiology of the 19th century**. *FEMS Microbiology Reviews* 2000, **24**:225–249.

[60] Govindjee, Shevela D: **Adventures with cyanobacteria: a personal perspective**. *frontiers in plant science* 2011, **2(28)**:1–17.

[61] Bergey DH, Buchanan RE, Gibbons NE: *Bergey's Manual of Determinative Bacteriology*. Williams & Wilkins Co. 1974.

[62] Kuetzing FT: *Tabulae Phycologicae - Abbildung der Tange (Band 1)*. Nordhausen. Auf Kosten des Verfassers 1845-1849.

[63] Naegeli CW: *Gattungen einzelliger Algen*. Friedrich Schulthess, Zuerich 1849.

[64] Castenholz RW: *Bergey's Manual of Systematic Bacteriology: The Archaea and the Deeply Branching and Phototropic Bacteria: Cyanobacteria*. Springer Verlag 2001.

[65] Rabenhorst L: *Deutschlands Kryptogamen-Flora - Erster Band Pilze*. Verlag von Eduard Kummer, Leipzig 1844.

[66] Rabenhorst L: *Deutschland's Kryptogamen-Flora - Zweiter Band , Dritte Abteilung: Leber-, Laubmoose und Farrn*. Verlag von Eduard Kummer, Leipzig 1848.

[67] Hauck F: *Kryptogamen-Flora - Die Meeresalgen*. Verlag von Eduard Kummer, Leipzig 1885.

[68] Holt JG, Murray RGE, Bergey DH, Krieg NA, Sneath PHA, Staley JT, Williams ST: *Bergey's Manual of Systematic Bacteriology (1. ed.)*. Williams & Wilkins, Baltimore, MD 1984-1989.

[69] Linne C: *Systema naturae per regna tria naturae, Volume 10th edition*. Stockholm: Laurentius Salvius. 1758.

[70] Oren A: **Naming Cyanophyta/Cyanobacteria – a bacteriologist's view**. *Fottea* 2011, **11(1)**:9–16.

[71] Copeland HF: *The Classification of lower organisms*. Pacific Books 1956.

[72] Engler A, Prantl K: *Die natuerlichen Pflanzenfamilien*. Verlag von Wilhelm Engelmann,Leipzig 1900.

[73] Haeckel E: *The wonders of life*. Harper & Brothers Publishers 1904.

[74] Lemmermann E: *Kryptogamenflora der Mark Brandenburg - Algen I (Schizophyceen, Flagellaten, Peridineen)*. Botanischer Verein der Provinz Brandenburg 1910.

[75] Buchanan RE: **Studies in the nomenclature and classification of the bacteria - II. The Primary Subdivisions of the Schizomycetes**. *J Bacteriol.* 1917, **2(2)**:155–164.

[76] Heller HH: **Phylogenetic Position of the Bacteria**. *Botanical Gazette* 1921, **72(6)**:390–396.

[77] Pringsheim EG: **The Relationship Between Bacteria and Myxophyceae**. *Bacteriological Reviews* 1949, **13**(2):47–98.

[78] Oren A: *Molecular Phylogeny of Microorganisms - Concepst About Phylogeny of Microorganisms*. Caister Academic Press, UK 2010.

[79] Satnier RY, VanNiel CB: **The main outlines of bacterial classification**. *J Bacteriol.* 1941, **42**:437–466.

[80] Copeland HF: **The Kingdoms of Organisms**. *The Quarterly Review of Biology* 1938, **13(4)**:383–420.

[81] Winogradsky S: **Sur la classification de bacteriales**. *Annales de L'institute Pasteur* 1952, **82(2)**:125–131.

[82] Stanier RY, vanNiel CB: **Concept of A Bacterium**. *Archiv Fur Mikrobiologie* 1962, **42**:17 – 35.

[83] Rippka R, Cohenbazire G: **The Cyanobacteriales - A Legitimate Order Based On the Type Strain Cyanobacterium-stanieri**. *Annales De Microbiologie* 1983, **B134**:21 – 36.

[84] Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY: **Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria**. *Journal of General Microbiology* 1979, **111**(MAR):1–61.

[85] Turner S, Pryer KM, Miao VPW, Palmer JD: **Investigating deep phylogenetic relationships among cyanobacteria and plastids by small submit rRNA sequence analysis**. *Journal of Eukaryotic Microbiology* 1999, **46**(4):327–338.

[86] Urbach E, Robertson DL, Crisholm SW: **Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation**. *Nature* 2002, **355**:267–270.

[87] Winston JE: *Describing species - Practical Taxonomic Procedure for Biologists*. Columbia Unversity Press, New York 1999.

[88] McNeill JC, Barrie FR, Burdet HM, Demoulin V, Hawksworth DL, Marhold K, Nicolson DH, Prado J, Silva PC, Skog JE, JWiersema HJM, Turland NJSotEC: *International Code of Botanical Nomenclature (Vienna Code)*. A.R.G. Gantner Verlag KG. 2005.

[89] Sneath PHA: **A short history of the Bacteriological Code**. *www.the-icsp.org/misc/Code_history.htm* 2003.

[90] Stanier RY, Sistrom WR, Hansen TA, Whitton BA, Castenholz RW, Pfennig N, Gorlenko VN, Kondratieva EN, Eimhjellen KE, Whittenbury R, Gherna RL, Trüper HG: **Proposal to Place the Nomenclature of the Cyanobacteria (Blue-Green Algae) Under the Rules of the International Code of Nomenclature of Bacteria**. *International Journal of Systematic and Microbiology* 1978, **28(2)**:335–336.

[91] Golubic S: **Cyanobacteria (Blue-Green Algae) under the Bacteriological Code? An Ecological Objection**. *Taxon* 1979, **28(4)**:387–389.

[92] Friedmann EI, Borowitzka LJ: **The Symposium on Taxonomic Concepts in Blue-Green Algae: Towards a Compromise with the Bacteriological Code?** *Taxon* 1982, **31(4)**:673–683.

[93] Oren A: **Cyanobacterial systematics and nomenclature as featured in the International Bulletin of Bacteriological Nomenclature and Taxonomy / International Journal of Systematic Bacteriology / International Journal of Systematic and Evolutionary Microbiology**. *International Journal of Systematic and Evolutionary Microbiology* 2011, **61**:10–15.

[94] Oren A: **A proposal for further integration of the cyanobacteria under the Bacteriological Code**. *International Journal of Systematic and Evolutionary Microbiology* 2004, **54**:1895–1902.

[95] Sagan L: **On Origin of Mitosing Cells**. *Journal of Theoretical Biology* 1967, **14**(3):225–274.

[96] Nelissen B, Vanderpeer Y, Wilmotte A, Dewachter R: **An Early Origin of Plastids Within the Cyanobacterial Divergence Is Suggested By Evolutionary Trees Based On Complete 16s Ribosomal-rna Sequences Rid D-4388-2009 Rid H-1686-2011**. *Molecular Biology and Evolution* 1995, **12**(6):1166–1173.

[97] Martin W, Kowallik KV: **Annotated English translation of Mereschkowsky's 1905 paper 'Uber Natur und Ursprung der Chromatophoren im Pflanzenreiche '**. *Eur. J. Phycol.* 1999, **34**:287–295.

[98] Gould SB, Waller RR, McFadden GI: **Plastid evolution**. *Annual Review of Plant Biology* 2008, **59**:491–517.

[99] Chan CX, Bhattacharya D: *Plastid Origin and Evolution. In: eLS*. John Wiley & Sons, Ltd: Chichester. 2011.

[100] Barghoorn ES, Tyler SA: **Microorganisms From Gunflint Chert - These Structurally Preserved Precambrian Fossils From Ontario Are Most Ancient Organisms Known**. *Science* 1965, **147**(3658):563–&.

[101] Cloud PE: **Significance of Gunflint (Precambrian) Microflora - Photosynthetic Oxygen May Have Had Important Local Effects Before Becoming A Major Atmospheric Gas**. *Science* 1965, **148**(3666):27–35.

[102] Schopf JW, Kvenvolden KA, S BE: **Amino Acids In Precambrian Sediments - An Assay**. *PNAS* 1968, **59**(2):639–646.

[103] Hofmann HJ: **Precambrian Microflora, Belcher Islands, Canada - Significance and Systematics**. *Journal of Paleontology* 1976, **50**(6):1040–1073.

[104] Golubic S, Campbell SE: **Analogous Microbial Forms In Recent Subaerial Habitats and In Precambrian Cherts - Gloeothece-coerulea Geitler and Eosynechococcus-moorei Hofmann**. *Precambrian Research* 1979, **8**(3-4):201–217.

[105] Schopf W: **Microfossils of the Early Archean Apex Chert: New Evidence of the Antiquity of life**. *Science* 1993, **260**:640–646.

[106] Schopf JW: **Disparate Rates, Differing Fates - Tempo and Mode of Evolution Changed From the Precambrian To the Phanerozoic**. *Proceedings of the National Academy of Sciences of the United States of America* 1994, **91**(15):6735–6742.

[107] Golubic S, Lee SJ: **Early cyanobacterial fossil record: preservation, palaeoenvironments and identification**. *European Journal of Phycology* 1999, **34**(4):339–348.

[108] Schopf JW: **Fossil evidence of Archaean life**. *Philosophical Transactions of the Royal Society B-biological Sciences* 2006, **361**(1470):869–885.

[109] Holland HD: **Volcanic gases, black smokers, and the Great Oxidation Event**. *Geochimica Et Cosmochimica Acta* 2002, **66**(21):3811–3826.

[110] Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ: **Dating the rise of atmospheric oxygen**. *Nature* 2004, **427**:117–120.

[111] Frei R, Gaucher C, Poulton SW, Canfield DE: **Fluctuations in Precambrian atmospheric oxygenation recorded by chromium isotopes**. *Nature* 2009, **461**(7261):250–254.

[112] Giovannoni SJ, Turner S, Olsen GJ, Barns S, Lane DJ, Pace NR: **Evolutionary Relationships Among Cyanobacteria And Green Chloroplasts**. *Journal of Bacteriology* 1988, **170**:3584–3592.

[113] Rippka R, Waterbury J, Cohenbazire G: **Cyanobacterium Which Lacks Thylakoids**. *Archives of Microbiology* 1974, **100**(4):419–436.

[114] Ishida T, Yokota A, Sugiyama J: **Phylogenetic relationships of filamentous cyanobacterial taxa inferred from 16S rRNA sequence divergence**. *Journal of General and Applied Microbiology* 1997, **43**(4):237–241.

[115] Honda D, Yokota A, Sugiyama J: **Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine Synechococcus strains**. *Journal Of Molecular Evolution* 1999, **48**(6):723–739.

[116] Ishida T, Watanabe MM, Sugiyama J, Yokota A: **Evidence for polyphyletic origin of the members of the orders of Oscillatoriales and Pleurocapsales as determined by 16S rDNA analysis**. *Fems Microbiology Letters* 2001, **201**:79–82.

[117] Wilmotte A, Neefs JM, Dewachter R: **Evolutionary Affiliation of the Marine Nitrogen-fixing Cyanobacterium Trichodesmium Sp Strain Nibb-1067, Derived By 16s-ribosomal Rna Sequence-analysis Rid H-1686-2011**. *Microbiology-uk* 1994, **140**:2159–2164.

[118] Sanchez-Baracaldo P, Hayes PK, Blank CE: **Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach**. *Geobiology* 2005, **3**(3):145–165.

[119] Swingley WD, Blankenship RE, Raymond J: **Integrating markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families**. *Molecular Biology and Evolution* 2008, **25**(4):643–654.

[120] Blank CE, Sánchez-Baracaldo P: **Timing of morphological and ecological innovations in the cyanobacteria - a key to understanding the rise in atmospheric oxygen**. *Geobiology* 2010, **8**:1–23.

[121] Larsson J, Nylander JAA, Bergman B: **Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits**. *Bmc Evolutionary Biology* 2011, **11**:187.

[122] Golubic S, Sergeev VN, Knoll AH: **Mesoproterozoic Archaeoellipsoides: Akinetes of heterocystous cyanobacteria**. *Lethaia* 1995, **28**(4):285–298.

[123] Amard B, BertrandSarfati J: **Microfossils in 2000 Ma old cherty stromatolites of the Franceville Group, Gabon**. *Precambrian Research* 1997, **81**(3-4):197–221.

[124] Zhang JZ: **Evolution by gene duplication: an update**. *Trends In Ecology & Evolution* 2003, **18**(6):292–298.

[125] Schrider DR, Hahn MW: **Gene copy-number polymorphism in nature**. *Proceedings of the Royal Society B-biological Sciences* 2010, **277**(1698):3213–3221.

[126] Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod HL, Cheverud JM, Ley TJ: **A high-resolution map of segmental DNA copy number variation in the mouse genome**. *Plos Genetics* 2007, **3**:e3.

[127] Springer NM, Ying K, Fu Y, Ji TM, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS: **Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content**. *Plos Genetics* 2009, **5**(11):e1000734.

[128] Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, Santos MAS: **Comparative genomics of wild type yeast strains unveils important genome diversity**. *Bmc Genomics* 2008, **9**:524.

[129] Beckmann JS, Estivill X, Antonarakis SE: **Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability**. *Nature Reviews Genetics* 2007, **8**(8):639–646.

[130] Perry GH: **The evolutionary significance of copy number variation in the human genome**. *Cytogenetic and Genome Research* 2008, **123**(1-4):283–287.

[131] Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC: **Diet and the evolution of human amylase gene copy number variation**. *Nature Genetics* 2007, **39**(10):1256–1260.

[132] Coenye T, Vandamme P: **Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes**. *RFEMS Microbiology Letters* 2003, **228**:45–49.

[133] Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Poles M, Brown SM, Sotero S, DeSantis T, Brodie E, Nelson K, Pei Z: **Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes**. *Applied and Environmental Microbiology* 2010, **76(12)**:3886–3897.

[134] Klappenbach JA, Dunbar JM, Schmidt TM: **rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria**. *Applied and Environmental Microbiology* 2000, **66(4)**:1328–1333.

[135] Tourova TP: **Copy number of ribosomal operons in prokaryotes and its effect on phylogenetic analyses**. *Microbiology* 2003, **72**(4):389–402.

[136] Liao DQ: **Gene conversion drives within genic sequences: Concerted evolution of ribosomal RNA genes in bacteria and archaea**. *Journal of Molecular Evolution* 2000, **51**(4):305–317.

[137] Rastogi R, Wu M, DasGupta I, Fox GE: **Visualization of ribosomal RNA operon copy number distribution**. *Bmc Microbiology* 2009, **9**:208.

[138] Maniloff J: **The minimal cell genome: "On being the right size"**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(19):10004–10006.

[139] Marais GAB, Calteau A, Tenaillon O: **Mutation rate and genome reduction in endosymbiotic and free-living bacteria**. *Genetica* 2008, **134**(2):205–210.

[140] Kuo CH, Moran NA, Ochman H: **The consequences of genetic drift for bacterial genome complexity**. *Genome Research* 2009, **19**(8):1450–1454.

[141] Litvaitis MK: **A molecular test of cyanobacterial phylogeny: inferences from constraint analyses**. *Hydrobiologia* 2002, **468**(1-3):135–145.

[142] Gugger MF, Hoffmann L: **Polyphyly of true branching cyanobacteria (stigonematales)**. *International Journal of Systematic and Evolutionary Microbiology* 2004, **54**:349–357.

[143] Tomitani A, Knoll AH, Cavanaugh CM, Ohno T: **The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(14):5442–5447.

[144] Fredriksson C, Bergman B: **Ultrastructural characterisation of cells specialised for nitrogen fixation in a non-heterocystous cyanobacterium, Trichodesmium spp.** *Protoplasma* 1997, **197**(1-2):76–85.

[145] Berman-Frank I, Lundgren P, Chen YB, Kupper H, Kolber Z, Bergman B, Falkowski P: **Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium Trichodesmium**. *Science* 2001, **294**(5546):1534–1537.

[146] El-Shehawy R, Lugomela C, Ernst A, Bergman B: **Diurnal expression of hetR and diazocyte development in the filamentous non-heterocystous cyanobacterium *Trichodesmium erythraeum***. *Microbiology-sgm* 2003, **149**:1139–1146.

[147] Ohki K, Taniuchi Y: **Detection of nitrogenase in individual cells of a natural population of trichodesmium using immunocytochemical methods for fluorescent cells**. *Journal of Oceanography* 2009, **65**(3):427–432.

[148] Zehr JP: **Nitrogen fixation by marine cyanobacteria**. *Trends In Microbiology* 2011, **19**(4):162–173.

[149] Engene N, Coates RC, Gerwick WH: **16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus Lyngbya**. *Journal of Phycology* 2010, **46**(3):591–601.

[150] Engene N, Gerwick WH: **Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes**. *Fottea* 2011, **11**:17–24.

[151] Noller HF, Woese CR: **Secondary Structure of 16S-ribosomal RNA**. *Science* 1981, **212**(4493):403–411.

[152] Olsen GJ, Woese CR: **Ribosomal-RNA - A Key To Phylogeny**. *Faseb Journal* 1993, **7**:113–123.

[153] Olivier A, Lee HY, Côté JC: **Study of the heterogeneity of 16S rRNA genes in γ-proteobacteria: Implications for phylogenetic analysis**. *Journal of General and Applied Microbiology* 2005, **51**:395–405.

[154] Aziz RK, Breitbart M, Edwards RA: **Transposases are the most abundant, most ubiquitous genes in nature RID B-2918-2009**. *Nucleic Acids Research* 2010, **38**(13):4207–4217.

[155] Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM: **Effect of temperature and light on growth of and photosynthesis by Synechococcus isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park**. *Applied and Environmental Microbiology* 2006, **72**:544–550.

[156] Steunou AS, Bhaya D, Bateson MM, Melendrez MC, Ward DM, Brecht E, Peters JW, Kuhl M, Grossman AR: **In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats RID A-1977-2009**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(7):2398–2403.

[157] Gupta R, Mathews D: **Signature proteins for the major clades of Cyanobacteria**. *BMC Evolutionary Biology* 2010, **10**:24.

[158] Criscuolo A, Gribaldo S: **Large-Scale Phylogenomic Analyses Indicate a Deep Origin of Primary Plastids within Cyanobacteria**. *Molecular Biology and Evolution* 2011, **28 (11)**:3019–3032.

[159] Schirrmeister BE, Antonelli A, Bagheri HC: **The origin of multicellularity in cyanobacteria**. *BMC Evol. Biol.* 2011, **11**:45.

[160] Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi C, Yamada M, Tabata S: **Complete genome structure of Gloeobacter violaceus PCC 7421, a cyanobacterium that lacks thylakoids**. *Dna Research* 2003, **10**(4):137–145.

[161] Lee ZM, Bussema C, Schmidt TM: **rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea**. *Nucleic Acid Res.* 2009, **37**:D489–D493.

[162] Adams DG: **Heterocyst formation in cyanobacteria**. *Current Opinion In Microbiology* 2000, **3**(6):618–624.

[163] Bjornsson L, Hugenholtz P, Tyson GW, Blackall LL: **Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal**. *Microbiology-sgm* 2002, **148**:2309–2318.

[164] Costello EK, Schmidt SK: **Microbial diversity in alpine tundra wet meadow soil: novel Chloroflexi from a cold, water-saturated environment**. *Environmental Microbiology* 2006, **8**(8):1471–1486.

[165] Nei M, Rogozin IB, Piontkivska H: **Purifying selection and birth-and-death evolution in the ubiquitin gene family**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(20):10866–10871.

[166] Sang T, Crawford DJ, Stuessy TF: **Documentation of Reticulate Evolution In Peonies (peonia) Using Internal Transcribed Spacer Sequences of Nuclear Ribosomal Dna - Implications For Biogeography and Concerted Evolution**. *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92**(15):6813–6817.

[167] Ganley ARD, Kobayashi T: **Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data**. *Genome Research* 2007, **17**(2):184–191.

[168] Santoyo G, Romero D: **Gene conversion and concerted evolution in bacterial genomes**. *Fems Microbiology Reviews* 2005, **29**(2):169–183.

[169] Simpson GG: *Tempo and Mode in Evolution*. Columbia University Press 1944.

[170] Schirrmeister BE, Anisimova M, Antonelli A, Bagheri HC: **Evolution of cyanobacterial morphotypes: Taxa required for improved phylogenomic approaches**. *Communicative and Integrative Biology* 2011, **4**:424–427.

[171] Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, Chisholm SW: **Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation**. *Nature* 2003, **424**(6952):1042–1047.

[172] Mazard SL, Fuller NJ, Orcutt KM, Bridle O, Scanlan DJ: **PCR analysis of the distribution of unicellular cyanobacterial diazotrophs in the Arabian Sea**. *Applied and Environmental Microbiology* 2004, **70**(12):7355–7364.

[173] Roth ACJ, Gonnet GH, Dessimoz C: **Algorithm of OMA for large-scale orthology inference**. *BMC Bioinformatics* 2008, **9**:518.

[174] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank**. *Nucleic Acids Research* 2008, **36**:D25–D30.

[175] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools**. *Nucleic Acids Research* 1997, **25**(24):4876–4882.

[176] Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models**. *Bioinformatics* 2003, **19**(12):1572–1574.

[177] Rambaut A, Drummond AJ: **Tracer v1.4 [online].** *available at: http://tree.bio.ed.ac.uk/software/tracer/* 2007.

[178] Rambaut A: **FigTree v1.3.1 [online]**. *available at: http://tree.bio.ed.ac.uk/software/figtree* 2009.

[179] Yang ZH: **PAML 4: Phylogenetic analysis by maximum likelihood**. *Molecular Biology and Evolution* 2007, **24**(8):1586–1591.

[180] Lemon J: **Plotrix: a package in the red light district of R**. *R-news* 2006, **6(4)**:8–12.

[181] Blankenship RE: *Molecular Mechanisms of Photosynthesis*. Blackwell Science, Oxford 2002.

[182] Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ: **The paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis**. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 2005, **102**(32):11131–11136.

[183] Allen JF, Martin W: **Evolutionary biology: Out of thin air**. *Nature* 2007, **445**:610–612.

[184] Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR: **Reassessing the first appearance of eukaryotes and cyanobacteria**. *Nature* 2008, **455**(7216):1101–4.

[185] Bonner J: **The Origin of Multicellularity**. *Integr.Biol.* 1998, **1**:28–36.

[186] Solari C, Kessler J, RE G: **Motility, mixing, and multicellularity**. *Genetic Programming and Evolvable Machines* 2007, **8**:115–129.

[187] Butterfield NJ: **Modes of pre-Ediacaran multicellularity**. *Precambrian Research* 2009, **173**(1-4):201–211.

[188] Giddings TJ, Staehelin L: **Observation of Mieroplasmodesmata in both Heterocyst-Forming and Non-Heterocyst Forming Filamentous Cyanobacteria by Freeze-Fracture Electron Microscopy**. *Archives of Microbiology* 1981, **129**:295–298.

[189] Flores E, Herrero A, Wolk CP, Maldener I: **Is the periplasm continuous in filamentous multicellular cyanobacteria?** *Trends In Microbiology* 2006, **14**(10):439–443.

[190] Mullineaux C, Mariscal V, Nenninger A, Khanum H, Herrero A, Flores E, Adams D: **Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria**. *European Molecular Biology Organization Journal* 2008, **27**:1299–1308.

[191] Flores E, Herrero A: **Compartmentalized function through cell differentiation in filamentous cyanobacteria**. *Nature Reviews Microbiology* 2010, **8**:39–50.

[192] Bonner J: *The evolution of complexity by means of natural selection*. Princeton University Press 1988.

[193] Sergeev VN: **The distribution of microfossil assemblages in Proterozoic rocks**. *Precambrian Research* 2009, **173**(1-4):212–222.

[194] Walsh MM: **Microfossils And Possible Microfossils From The Early Archean Onverwacht Group, Barberton Mountain Land, South-Africa**. *Precambrian Research* 1992, **54**(2-4):271–293.

[195] Tice MM, Lowe DR: **Photosynthetic microbial mats in the 3,416-Myr-old ocean**. *Nature* 2004, **431**:549–552.

[196] Westall F, deVries ST, Nijman W, Rouchon V, Orberrger B, Pearson V, Watson J, Verchovsky A, Rouzaud JN, Marchesin D, Severine A: **The 3.466 Ga "Kitty's Gap Chert" an early Archean microbial ecosystem**. *Geological Society of America Special Papers* 2006, **405**:105–131.

[197] Wacey D: *Early Life on Earth: A practical Guide*. Springer 2009.

[198] Schopf JW, Kudryavtsev AB, Czaja AD, Tripathi AB: **Evidence of archean life: Stromatolites and microfossils**. *Precambrian Research* 2007, **158(3-4)**:141–155.

[199] Brasier M, Green O, Lindsay J, Steele A: **Earth's oldest (similar to 3.5 Ga) fossils and the 'Early Eden hypothesis': Questioning the evidence**. *Origins Of Life And Evolution Of The Biosphere* 2004, **34**(1-2):257–269.

[200] Brasier M, McLoughlin N, Green O, Wacey D: **A fresh look at the fossil evidence for early Archaean cellular life RID D-4080-2011**. *Philosophical Transactions of the Royal Society B-biological Sciences* 2006, **361**(1470):887–902.

[201] Garcia-Pichel F, Lopez-Cortes A, Nubel U: **Phylogenetic and morphological diversity of cyanobacteria in soil desert crusts from the Colorado Plateau**. *Applied And Environmental Microbiology* 2001, **67**(4):1902–1910.

[202] Doolittle WF: **Phylogenetic classification and the universal tree**. *Science* 1999, **284**(5423):2124–2128.

[203] Suchard MA: **Stochastic models for horizontal gene transfer: Taking a random walk through tree space**. *Genetics* 2005, **170**:419–431.

[204] Hamel L, Zhaxybayeva O, Gogarten J: **PentaPlot: A software tool for the illustration of genome mosaicism**. *BMC Bioinformatics* 2005, **6**:139.

[205] Hess WR, Rocap G, Ting CS, Larimer F, Stilwagen S, Lamerdin J, Chisholm SW: **The photosynthetic apparatus of Prochlorococcus: Insights through comparative genomics**. *Photosynthesis Research* 2001, **70**:53–71.

[206] Robertson BR, Tezuka N, Watanabe MM: **Phylogenetic analyses of Synechococcus strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content**. *International Journal Of Systematic And Evolutionary Microbiology* 2001, **51**:861–871.

[207] Ernst A, Becker S, Wollenzien UIA, Postius C: **Ecosystem-dependent adaptive radiations of picocyanobacteria inferred from 16S rRNA and ITS-1 sequence analysis**. *Microbiology-Sgm* 2003, **149**:217–228.

[208] Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C, Ferriera S, Johnson J, Post AF, Hess WR, Partensky F: **Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria**. *Genome Biology* 2008, **9**(5):R90.

[209] Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F: **Ecological Genomics of Marine Picocyanobacteria**. *Microbiology And Molecular Biology Reviews* 2009, **73**(2):249–+.

[210] Brochier C, Bapteste E, Moreira D, Philippe H: **Eubacterial phylogeny based on translational apparatus proteins**. *Trends in Genetics* 2002, **18**:1–4.

[211] Battistuzzi FU, Feijao A, Hedges SB: **A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land**. *Bmc Evolutionary Biology* 2004, **4**:44.

[212] Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P: **Toward automatic reconstruction of a highly resolved tree of life RID B-3300-2008 RID B-8886-2011**. *Science* 2006, **311**(5765):1283–1287.

[213] Battistuzzi FU, Hedges SB: **A Major Clade of Prokaryotes with Ancient Adaptations to Life on Land**. *Mol. Biol. Evol.* 2009, **26(2)**:335–343.

[214] Delwiche CF, Kuhsel M, PalmerR JD: **Phylogenetic Analysis Of Tufa Sequences Indicates A Cyanobacterial Origin Of All Plastids**. *Molecular Phylogenetics And Evolution* 1995, **4**(2):110–128.

[215] Seo P, Yokota A: **The phylogenetic relationships of cyanobacteria inferred from 16SrRNA, gyrB, rpoC1 and rpoD1 gene sequences**. *The Journal of General and Applied Microbiology* 2003, **49**:191—203.

[216] Mangels D, Kruip J, Berry S, Rogner M, Boekema EJ, Koenig F: **Photosystem I from the unusual cyanobacterium Gloeobacter violaceus**. *Photosynthesis Research* 2002, **72**:307–319.

[217] Siefert JL, Fox GE: **Phylogenetic mapping of bacterial morphology**. *Microbiology-uk* 1998, **144**:2803–2808.

[218] Kremer B: **Mat-forming coccoid cyanobacteria from early Silurian marine deposits of Sudetes, Poland**. *Acta Palaeontologica Polonica* 2006, **51**:143–154.

[219] Maddison W, Maddison D: **Mesquite: a modular system for evolutionary analysis.** *Version 2.71 http://mesquiteproject.org* 2009.

[220] Erixon P, Svennblad B, Britton T, Oxelman B: **Reliability of Bayesian Posterior Probabilities and Bootstrap Frequencies in Phylogenetics**. *Syst. Biol.* 2003, **52(5)**:665–673.

[221] Huelsenbeck JP, Rannala B: **Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models**. *Systematic Biology* 2004, **53**(6):904–913.

[222] Alfaro ME, Holder MT: **The Posterior and the Prior in Bayesian Phylogenetics**. *Annu. Rev. Ecol. Evol. Syst.* 2006, **37**:19–42.

[223] Stucken K, John U, Cembella A, Murillo AA, Soto-Liebe K, Fuentes-Valdes JJ, Friedel M, Plominsky AM, Vasquez M, Glockner G: **The Smallest Known Genomes of Multicellular and Toxic Cyanobacteria: Comparison, Minimal Gene Sets for Linked Traits and the Evolutionary Implications**. *Plos One* 2010, **5**(2):e9235.

[224] Knoll AH, Javaux EJ, Hewitt D, Cohen P: **Eukaryotic organisms in Proterozoic oceans**. *Philosophical Transactions of the Royal Society B-biological Sciences* 2006, **361**(1470):1023–1038.

[225] Buss LW: *The Evolution of Individuality*. Princeton University Press 1987.

[226] Maynard Smith J, Szathmary E: *The Major Transitions in Evolution*. W.H.Freeman and Spektrum 1995.

[227] Buss LW: **Slime molds, ascidians, and the utility of evolutionary theory**. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 1999, **96**(16):8801–8803.

[228] Michod R: *Darwinian Dynamics-Evolutionary Transitions in Fitness and Individuality*. Princeton University Press 1999.

[229] Tuomi J: **Genetic heterogeneity within organisms and the evolution of individuality**. *Journal Of Evolutionary Biology* 2004, **17**(6):1182–1183.

[230] Cameron S, Mardulyn P: **Multiple Molecular Data Sets Suggest Independent Origins of Highly Eusocial Behavior in Bees (Hymenoptera: Apinae)**. *Syst. Biol.* 2001, **50**:194–214.

[231] Danforth BN: **Evolution of sociality in a primitively eusocial lineage of bees**. *PNAS* 2002, **99**:286–290.

[232] Carroll SB: **Chance and necessity: the evolution of morphological complexity and diversity**. *Nature* 2001, **409**:1102–1109.

[233] King N: **The Unicellular Ancestry of Animal Development**. *Developmental Cell* 2004, **7**:313–325.

[234] Grosberg R, Strathmann R: **The Evolution of Multicellularity: A Minor Major Transition?** *Annu.Rev.Ecol.Evol.Syst.* 2007, **38**:621–54.

[235] Danforth BN, Conway L, Shuqing J: **Phylogeny of Eusocial Lasioglossum Reveals Multiple Losses of Eusociality Within a Primitively Eusocial Clade of Bees (Hymenoptera:Halictidae)**. *Syst.Biol.* 2003, **52**:23–36.

[236] Chapman H, Houliston GJ, Robson B, Iline I: **A case of reversal: The evolution and maintenance of sexuals from parthenogenetic clones in Hieracium pilosella**. *International Journal of Plant Science* 2003, **164**:719–728.

[237] Whiting MF, Bradler S, Maxwell T: **Loss and recovery of wings in stick insects**. *Nature* 2003, **421**:264–267.

[238] Domes K, Norton RA, Maraun M, Scheu S: **Reevolution of sexuality breaks Dollo's law**. *PNAS* 2007, **104**:7139–7144.

[239] Zufall RA, Rausher MD: **Genetic changes associated with floral adaptation restrict future evolutionary potential**. *Nature* 2004, **428**:847–850.

[240] Igic B, Bohs L, Kohn J: **Ancient polymorphism reveals unidirectional breeding system shifts**. *PNAS* 2006, **103**:1359–1363.

[241] Bull JJ, Charnov EL: **On Irreversible Evolution**. *Evolution* 1985, **39**:1149–1155.

[242] Brasier M, Green O, Jephcoat A, Kleppe A, Van Kranendonk J, Lindsay J, Steele A, Grassineau N: **Questioning the evidence for Earth's oldest fossils**. *Nature* 2002, **416**:76–81.

[243] Noffke N, Eriksson KA, Hazen RM, Simpson EL: **A new window into Early Archean life: Microbial mats in Earth's oldest siliciclastic tidal deposits (3.2 Ga Moodies Group, South Africa)**. *Geology* 2006, **34**(4):253–256.

[244] Sugitani K, Grey K, Allwood A, Nagaoka T, Mimura K, Minami M, Marshall CP, Van Kranendonk MJ, Walter MR: **Diverse microstructures from Archaean chert from the mount goldsworthy-mount grant area, Pilbara Craton, Western Australia: Microfossils, dubiofossils, or pseudofossils?** *Precambrian Research* 2007, **158**:228–262.

[245] Glikson M, Duck LJ, Golding A S Dand Hofmann, Bolhar R, Webb R, Baiano J, Sly L: **Microbial remains in some earliest Earth rocks: Comparison with a potential modern analogue**. *Precambrian Research* 2008, **164**:187–200.

[246] Wacey D, Kilburn MR, McLoughlin N, Parnell J, Stoake CA, Grovenor CRM, Brasier MD: **Use of NanoSIMS in the search for early life on Earth: ambient inclusion trails in a c. 3400 Ma sandstone**. *Journal of the Geological Society* 2008, **1**:43–53.

[247] Klein C, Beukes N, Schopf J: **Filamentous microfossils in the early proterozoic transvaal supergroup: their morphology, significance, and paleoenvironmental setting**. *Precambrian Research* 1987, **36**:81–94.

[248] Altermann W, Schopf J: **Microfossils from the Neoarchean Campbell Group, Griqualand West Sequence of the Transvaal Supergroup, and their paleoenvironmental and evolutionary implications**. *Precambrian Research* 1995, **75**:65–90.

[249] Schopf J, Packer B: **Early Archean (3.3-Billion to 3.5-Billion-Year-Old) Microfossils from Warrawoona Group, Australia**. *Science* 1987, **237**:70–73.

[250] Knoll A, Barghoorn E: **Archaean Microfossils Showing Cell Division from the Swaziland System of South Africa**. *Science* 1977, **198**:396–398.

[251] Schopf J, Barghoorn E: **Alga-Like Fossils from the Early Precambrian of South Africa**. *Science* 1967, **156**:506–512.

[252] Rosing M, Bird D, Sleep N, Glassley W, Albarede F: **The rise of continents - An essay on the geologic consequences of photosynthesis**. *Science direct* 2006, **232**:99–113.

[253] Wilde SA, Valley JW, Peck WH, Graham CM: **Evidence from detrital zircons for the existence of continental crust and ceans on the Earth 4.4 Gyr ago**. *Nature* 2001, **409**:175–178.

[254] Evans D, Beukes N, Kirschvink J: **Low-latitude glaciation in the Palaeoproterozoic era**. *Nature* 1997, **386**:262–266.

[255] Rossetti V, Schirrmeister BE, Bernasconi MV, Bagheri HC: **The evolutionary path to terminal differentiation and division of labor in cyanobacteria RID A-9703-2010**. *Journal of Theoretical Biology* 2010, **262**:23–34.

[256] Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C, Garvin J, Buick R: **A whiff of oxygen before the Great Oxidation Event? RID A-5572-2010**. *Science* 2007, **317**(5846):1903–1906.

[257] Butterfield NJ: **Macroevolution and macroecology through deep time**. *Palaeontology* 2007, **50**:41–55.

[258] Damuth J: **Evolution: Tempo and Mode**. *Encyclopedia of Life Science* 2001, :1–7.

[259] Stanley SM: **Ecological Theory For Sudden Origin Of Multicellular Life In Late Precambrian - (Adaptive Radiation-Cambrian-Evolution-Paleontology-Predation)**. *Proceedings Of The National Academy Of Sciences Of The United States Of America* 1973, **70**(5):1486–1489.

[260] Boraas ME, Seale DB, Boxhorn JE: **Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity**. *Evolutionary Ecology* 1998, **12**(2):153–164.

[261] Kaiser D: **Building a multicellular organism**. *Annual Review Of Genetics* 2001, **35**:103–123.

[262] Tannenbaum E: **Selective advantage for multicellular replicative strategies: A two-cell example**. *Physical Review E* 2006, **73**:010904.

[263] Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids res.* 2002, **30**:3059–3066.

[264] Miller M, Holder M, Vos R, Midford P, Liebowitz T, Chan L, Hoover P, Warnow T: **The CIPRES Portals. CIPRES.** *http://www.phylo.org/sub_sections/portal* 2009.

[265] Hall T: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/97/NT**. *Oxford University Press* 1999, **41**:95–98.

[266] Stamatakis A: **RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models RID B-8740-2009**. *Bioinformatics* 2006, **22**(21):2688–2690.

[267] Lanave C, Preparata G, Saccone C, Serio G: **A New Method For Calculating Evolutionary Substitution Rates**. *Journal of Molecular Evolution* 1984, **20**:86–93.

[268] Rodriguez F, Oliver JL, Marin A, Medina JR: **The General Stochastic-Model Of Nucleotide Substitution**. *Journal Of Theoretical Biology* 1990, **142**(4):485–501.

[269] Xia X, Xie Z: **DAMBE: Data analysis in molecular biology and evolution.** *Journal of Heredity* 2001, **92**:371–373.

[270] Xia X, Xie Z, Salemi M, Chen M, Wang Y: **An index of substitution saturation and its application**. *Molecular Phylogenetics and Evolution* 2003, **26**:1–7.

[271] Zwickl DJ: **Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.** *PhD-Dissertation* 2006, **The University of Texas at Austin**.

[272] Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution**. *Bioinformatics* 1998, **14**(9):817–818.

[273] Sukumaran J: **SumTrees: Summarization of Split Support on Phylogenetic Trees. Version 1.0.2**. *DendroPy Phylogenetic Computation Library Version 2.0.3* 2008.

[274] Wilgenbusch J, Warren D, Swofford D: **AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference**. *http://ceb.csit.fsu.edu/awty.* 2004.

[275] Pagel M, Meade A, Barker D: **Bayesian estimation of ancestral character states on phylogenies**. *Systematic Biology* 2004, **53**(5):673–684.

[276] Pardi F, Goldman N: **Species choice for comparative genomics: Being greedy works**. *Plos Genetics* 2005, **1**(6):e71.

[277] Delsuc F, Brinkmann H, Philippe H: **Phylogenomics and the reconstruction of the tree of life**. *Nature Reviews Genetics* 2005, **6**(5):361–375.

[278] Doolittle WF: **Uprooting the tree of life**. *Scientific American* 2000, **282**(2):90–95.

[279] Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M: **Parallel adaptations to high temperatures in the Archaean eon**. *Nature* 2008, **456**(7224):942–U74.

[280] Awramik SM, Barghoorn ES: **Gunflint Microbiota**. *Precambrian Research* 1977, **5**(2):121–142.

[281] Sergeev VN, Gerasimenko LM, Zavarzin GA: **The Proterozoic history and present state of cyanobacteria**. *Microbiology* 2002, **71**(6):623–637.

[282] Benton MJ: *The quality of the fossil record in Telling the Evolutionary Time: Molecular Clocks and the Fossil Record*. 66, Crc Press, Boca Raton, Florida 2003.

[283] Reisz RR, Muller J: **Molecular timescales and the fossil record: a paleontological perspective**. *Trends In Genetics* 2004, **20**(5):237–241.

[284] Donoghue PCJ, Benton MJ: **Rocks and clocks: calibrating the Tree of Life using fossils and molecules RID A-3873-2008 RID A-5639-2008**. *Trends In Ecology & Evolution* 2007, **22**(8):424–431.

[285] Drummond AJ, Ho SYW, Phillips MJ, Rambaut A: **Relaxed phylogenetics and dating with confidence RID A-3209-2010 RID A-8417-2008 RID B-2481-2009**. *Plos Biology* 2006, **4**(5):e88.

[286] Rossetti V, Filippini M, Svercel M, Barbour AD, Bagheri HC: **Emergent multicellular life cycles in filamentous bacteria owing to density-dependent population dynamics RID A-9728-2010**. *Journal of the Royal Society Interface* 2011, **8**(65):1772–1784.

[287] Ratcliff WC, Denison RF, Borrello M, Travisano M: **Experimental evolution of multicellularity**. *PNAS* 2011, :1–6.

[288] Koschwanez JH, Foster KR, Murray AW: **Sucrose Utilization in Budding Yeast as a Model for the Origin of Undifferentiated Multicellularity**. *Plos Biology* 2011, **9**(8):e1001122.

[289] Nisbet EG, Sleep NH: **The habitat and nature of early life**. *Nature* 2001, **409**(6823):1083–1091.

[290] Mojzsis SJ: **Early Earth leftover lithosphere**. *Nature Geoscience* 2010, **3**(3):148–149.

[291] Sleep NH, Zahnle KJ, Kasting JF, Morowitz HJ: **Annihilation of Ecosystems By Large Asteroid Impacts On the Early Earth**. *Nature* 1989, **342**(6246):139–142.

[292] Knoll AH: *Palynology: Principles and Applications - Archean and Proterozoic paleontology*. American Association of Stratigraphic Palynologists Press 1996.

[293] Kazmierczak J, Altermann W: **Neoarchean biomineralization by benthic cyanobacteria RID B-4336-2008**. *Science* 2002, **298**(5602):2351–2351.

[294] Schopf JW: *Encyclopedia of Microbiology - Paleontology, Microbial*. Amsterdam: Elsevier 2009.

[295] Brocks JJ, Logan GA, Buick R, Summons RE: **Archean molecular fossils and the early rise of eukaryotes**. *Science* 1999, **285**(5430):1033–1036.

[296] Summons RE, Jahnke LL, Hope JM, Logan GA: **2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis**. *Nature* 1999, **400**(6744):554–557.

[297] Lyons TW, Reinhard CT: **Earth Science- Sea change for the rise of oxygen**. *Nature* 2011, **478**(7368):194–195.

[298] Gaillard F, Scaillet B, Arndt NT: **Atmospheric oxygenation caused by a change in volcanic degassing pressure**. *Nature* 2011, **478**(7368):229–U112.

[299] Waldbauer JR, Sherman LS, Sumner DY, Summons RE: **Late Archean molecular fossils from the Transvaal Supergroup record the antiquity of microbial diversity and aerobiosis RID E-8744-2011**. *Precambrian Research* 2009, **169**(1-4):28–47.

[300] Bilofsky H, Burks C: **The genbank (r) genetic sequence data bank.** *Nuc. Acid Res.* 1988, **16**:1861–1864.

[301] Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic Acids Research* 2004, **32**(5):1792–1797.

[302] Anderson CL, Strope CL, Moriyama EN: **SuiteMSA: visual tools for multiple sequence alignment comparison and molecular sequence simulation**. *Bmc Bioinformatics* 2011, **12**:184.

[303] Akaike H: **New Look At Statistical-model Identification**. *Ieee Transactions On Automatic Control* 1974, **AC19**(6):716–723.

[304] Hurvich CM, Tsai CL: **Regression and Time-series Model Selection In Small Samples**. *Biometrika* 1989, **76**(2):297–307.

[305] Posada D: **jModelTest: Phylogenetic model averaging RID C-4502-2008**. *Molecular Biology and Evolution* 2008, **25**(7):1253–1256.

[306] Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees**. *Bmc Evolutionary Biology* 2007, **7**:214.

[307] Ferris MJ, RuffRoberts AL, Kopczynski ED, Bateson MM, Ward DM: **Enrichment culture and microscopy conceal diverse thermophilic Synechococcus populations in a single hot spring microbial mat habitat**. *Applied and Environmental Microbiology* 1996, **62**(3):1045–1050.

[308] Sukumaran J, Holder MT: **DendroPy: a Python library for phylogenetic computing**. *Bioinformatics* 2010, **26**(12):1569–1571.

[309] Ho SYW, Phillips MJ: **Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times RID A-8417-2008**. *Systematic Biology* 2009, **58**(3):367–380.

[310] Horodyski RJ, Donaldson JA: **Micro-fossils From the Middle Proterozoic Dismal Lakes Group, Arctic Canada**. *Precambrian Research* 1980, **11**(2):125–159.

[311] Sergeev VN, Knoll AH, Grotzinger JP: **Paleobiology of the Mesoproterozoic Billyakh Group, Anabar Uplift, Northern Siberia**. *Journal of Paleontology* 1995, **69**:1–37.

[312] Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ: **Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates**. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(32):13410–13414.

Additional Files

Additional Files to chapter I: Gene copy number variation and its significance in cyanobacterial phylogeny.

Additional File 1.1 - Identified gene copies (following pages)

The sheet contains Information on 41 gene copies and their presence in 22 cyanobacterial species. Amino acid sequences of the coded proteins exhibit 98% similarity within a genome and 50% across species.

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
1 1196Cyanothece sp PCC8801 protein	CP001287.1 ACK65303.1	gene=PCC8801 1237	protein=photosystem q(b) protein	protein id=ACK65303.1	location=1288606..1289676
1 1946Cyanothece sp PCC8801 protein	CP001287.1 ACK66053.1	gene=PCC8801 2016	protein=photosystem q(b) protein	protein id=ACK66053.1	location=2090110..2091180
2 1439Synechococcus sp RCC307 protein	CT978603.1 CAK28343.1	gene=psbA	protein=Photosystem II protein D1	protein id=CAK28343.1	location=complement(1279858..1280937)
2 2008Synechococcus sp RCC307 protein	CT978603.1 CAK28912.1	gene=psbA	protein=Photosystem II protein D1	protein id=CAK28912.1	location=complement(1742058..1743137)
2 2182Synechococcus sp RCC307 protein	CT978603.1 CAK29086.1	gene=psbA	protein=Photosystem II protein D1	protein id=CAK29086.1	location=1894037..1895116
5 1594Anabaena variabilis ATCC29413 protein	CP000117.1 ABA21220.1	gene=Ava 1597	protein=Photosystem II reaction centre protein PsbA/D1	protein id=ABA21220.1	location=complement(1971884..1972966)
5 2449Anabaena variabilis ATCC29413 protein	CP000117.1 ABA22075.1	gene=Ava 2460	protein=Photosystem II reaction centre protein PsbA/D1	protein id=ABA22075.1	location=complement(3053603..3054685)
5 3533Anabaena variabilis ATCC29413 protein	CP000117.1 ABA23159.1	gene=Ava 3553	protein=Photosystem II reaction centre protein PsbA/D1	protein id=ABA23159.1	location=complement(4422987..4424069)
6 4108Arthrospira platensis NIES-39 protein	AP011615.1 BAI91950.1	gene=psbAIV	protein=photosystem II reaction center D1 protein	protein id=BAI91950.1	location=complement(4171577..4172656)
6 4113Arthrospira platensis NIES-39 protein	AP011615.1 BAI91955.1	gene=psbAIII	protein=photosystem II reaction center D1 protein	protein id=BAI91955.1	location=complement(4178243..4179322)
6 6481Arthrospira platensis NIES-39 protein	AP011615.1 BAI94323.1	gene=psbAII	protein=photosystem II reaction center D1 protein	protein id=BAI94323.1	location=complement(6629459..6630538)
7 8Synechocystis sp PCC6803 protein	BA000022.2 BAA16586.1	gene=psbA2	protein=photosystem II D1 protein	protein id=BAA16586.1	location=7229..8311
7 1651Synechocystis sp PCC6803 protein	BA000022.2 BAA18230.1	gene=psbA3	protein=photosystem II D1 protein	protein id=BAA18230.1	location=complement(1818915..1819997)
8 779Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88720.1	gene=psbA	protein=photosystem II protein D1	protein id=BAC88720.1	location=832184..833266
8 2322Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC90263.1	gene=psbA	protein=photosystem II protein D1	protein id=BAC90263.1	location=2486614..2487696
8 3143Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC91085.1	gene=psbA	protein=photosystem II protein D1	protein id=BAC91085.1	location=complement(3348383..3349465)
9 165Cyanothece sp PCC7424 protein	CP001291.1 ACK68639.1	gene=PCC7424 0169	protein=photosystem q(b) protein	protein id=ACK68639.1	location=186554..187624
9 1028Cyanothece sp PCC7424 protein	CP001291.1 ACK69503.1	gene=PCC7424 1049	protein=photosystem q(b) protein	protein id=ACK69503.1	location=1150654..1151724
9 2240Cyanothece sp PCC7424 protein	CP001291.1 ACK70715.1	gene=PCC7424 2293	protein=photosystem q(b) protein	protein id=ACK70715.1	location=complement(2562885..2563955)
10 405Prochlorococcus marinus str MIT9303 protein	CP000554.1 ABM77161.1	gene=psbA	protein=Photosystem II PsbA protein (D1)	protein id=ABM77161.1	location=406385..407461
10 1853Prochlorococcus marinus str MIT9303 protein	CP000554.1 ABM78610.1	gene=P9303 18681	protein=Hypothetical protein	protein id=ABM78610.1	location=1629607..1630683
12 3568Nostoc sp PCC7120 protein	BA000019.2 BAB75271.1	gene=psbAIV	protein=photosystem II protein D1	protein id=BAB75271.1	location=complement(4315437..4316519)
12 3723Nostoc sp PCC7120 protein	BA000019.2 BAB75426.1	gene=psbAII	protein=photosystem II protein D1	protein id=BAB75426.1	location=4499998..4501080
12 4587Nostoc sp PCC7120 protein	BA000019.2 BAB76291.1	gene=psbAIII	protein=photosystem II protein D1	protein id=BAB76291.1	location=5489831..5490913
14 1021Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00844.1	gene=psbA1	protein=photosystem II D1 protein	protein id=BAG00844.1	location=complement(877533..878615)
14 1037Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00860.1	gene=psbA2	protein=photosystem II D1 protein	protein id=BAG00860.1	location=891585..892667
14 1050Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00873.1	gene=psbA3	protein=photosystem II D1 protein	protein id=BAG00873.1	location=900555..901637
14 1079Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00902.1	gene=psbA4	protein=photosystem II D1 protein	protein id=BAG00902.1	location=925990..927072
14 5813Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG05636.1	gene=psbA5	protein=photosystem II D1 protein	protein id=BAG05636.1	location=5351373..5352455
15 584Nostoc azollae 0708 protein	CP002059.1 ADI63253.1	gene=Aazo 0830	protein=photosystem q(b) protein	protein id=ADI63253.1	location=879395..880477
15 1149Nostoc azollae 0708 protein	CP002059.1 ADI63818.1	gene=Aazo 1652	protein=photosystem q(b) protein	protein id=ADI63818.1	location=1720331..1721413
16 2089Acharyochloris marina MBIC11017 protein	CP000828.1 ABW27180.1	gene=psbA	protein=photosystem II D1 protein PsbA	protein id=ABW27180.1	location=complement(2166714..2167796)
16 2799Acharyochloris marina MBIC11017 protein	CP000828.1 ABW27888.1	gene=psbA	protein=photosystem II D1 protein PsbA	protein id=ABW27888.1	location=complement(2928273..2929355)
17 2057Nostoc punctiforme PCC73102 protein	CP001037.1 ACC80861.1	gene=Npun R2273	protein=photosystem q(b) protein	protein id=ACC80861.1	location=complement(2808705..2809787)
17 3144Nostoc punctiforme PCC73102 protein	CP001037.1 ACC81948.1	gene=Npun F3544	protein=photosystem q(b) protein	protein id=ACC81948.1	location=4451534..4452616
17 4716Nostoc punctiforme PCC73102 protein	CP001037.1 ACC83520.1	gene=Npun R5188	protein=photosystem q(b) protein	protein id=ACC83520.1	location=complement(6424030..6425112)
18 166Synechococcus elongatus PCC6301 protein	AP008231.1 BAD78356.1	gene=psbAII	protein=photosystem II D1 protein	protein id=BAD78356.1	location=185180..186262
18 647Synechococcus elongatus PCC6301 protein	AP008231.1 BAD78837.1	gene=psbAIII	protein=photosystem II D1 protein	protein id=BAD78837.1	location=complement(716582..717664)
19 779Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88720.1	gene=psbA	protein=photosystem II protein D1	protein id=BAC88720.1	location=832184..833266
19 2322Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90263.1	gene=psbA	protein=photosystem II protein D1	protein id=BAC90263.1	location=2486614..2487696
19 3143Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC91085.1	gene=psbA	protein=photosystem II protein D1	protein id=BAC91085.1	location=complement(3348383..3349465)

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
20 143Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG49673.1	gene=Tery 0182	protein=photosystem q(b) protein	protein id=ABG49673.1	location=269811..270872
20 144Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG49674.1	gene=Tery 0183	protein=photosystem q(b) protein	protein id=ABG49674.1	location=271328..272389
20 4186Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG53717.1	gene=Tery 4763	protein=photosystem q(b) protein	protein id=ABG53717.1	location=complement(731951..731701 2)
21 3665synechococcus sp PCC7803 protein	CT971583.1 CAK22792.1	gene=psbA	protein=Photosystem II protein D1	protein id=CAK22792.1	location=complement(384540..385619)
21 7905synechococcus sp PCC7803 protein	CT971583.1 CAK23216.1	gene=psbA	protein=Photosystem II protein D1	protein id=CAK23216.1	location=776911..777990
21 20845synechococcus sp PCC7803 protein	CT971583.1 CAK24510.1	gene=psbA	protein=Photosystem II protein D1	protein id=CAK24510.1	location=complement(1914146..1915225)
22 1535synechococcus sp PCC7002 protein	CP000951.1 ACA98171.1	gene=psbA-II	protein=photosystem II D1 subunit PsbA-II (Qb protein)	protein id=ACA98171.1	location=complement(161801..162847)
22 1391Synechococcus sp PCC7002 protein	CP000951.1 ACA99409.1	gene=psbA	protein=photosystem q(b) protein	protein id=ACA99409.1	location=1489906..1490988
1 56Cyanothece sp PCC8801 protein	CP001287.1 ACK64163.1	gene=PCC8801 0056	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ACK64163.1	location=complement(58944..60002)
1 1891Cyanothece sp PCC8801 protein	CP001287.1 ACK65998.1	gene=PCC8801 1960	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ACK65998.1	location=2036321..2037379
2 2585Synechococcus sp RCC307 protein	CT978603.1 CAK27162.1	gene=psbD	protein=Photosystem II D2 protein	protein id=CAK27162.1	location=256888..257946
2 16945synechococcus sp RCC307 protein	CT978603.1 CAK28598.1	gene=psbD	protein=Photosystem II D2 protein	protein id=CAK28598.1	location=1479498..1480556
3 4547thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08007.1	gene=psbD2	protein=photosystem II reaction center D2 protein	protein id=BAC08007.1	location=456131..457189
3 1628Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC09182.1	gene=psbD1	protein=photosystem II reaction center D2 protein	protein id=BAC09182.1	location=1702500..1703558
5 1240Anabaena variabilis ATCC29413 protein	CP000117.1 ABA20866.1	gene=Ava 1242	protein=Photosystem II reaction centre protein PsbD/D2	protein id=ABA20866.1	location=1529297..1530352
5 2501Anabaena variabilis ATCC29413 protein	CP000117.1 ABA22127.1	gene=Ava 2512	protein=Photosystem II reaction centre protein PsbD/D2	protein id=ABA22127.1	location=complement(3106302..3107357)
6 1954Arthrospira platensis NIES-39 protein	AP011615.1 BAI89796.1	gene=psbDII	protein=photosystem II reaction center D2 protein	protein id=BAI89796.1	location=1968930..1969988
6 3148Arthrospira platensis NIES-39 protein	AP011615.1 BAI90990.1	gene=psbD1	protein=photosystem II reaction center D2 protein	protein id=BAI90990.1	location=3179356..3180414
7 12215synechocystis sp PCC6803 protein	BA000022.2 BAA17800.1	gene=psbD	protein=photosystem II D2 protein	protein id=BAA17800.1	location=complement(1348840..1349898)
7 28815synechocystis sp PCC6803 protein	BA000022.2 BAA10851.1	gene=psbD2	protein=photosystem II D2 protein	protein id=BAA10851.1	location=3229780..3230838
9 566Cyanothece sp PCC7424 protein	CP001291.1 ACK69040.1	gene=PCC7424 0579	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ACK69040.1	location=complement(636371..637429)
9 2902Cyanothece sp PCC7424 protein	CP001291.1 ACK71377.1	gene=PCC7424 2974	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ACK71377.1	location=complement(3317859..3318917)
12 4285Nostoc sp PCC7120 protein	BA000019.2 BAB75989.1	gene=psbD	protein=photosystem II protein D2	protein id=BAB75989.1	location=5143731..5144786
12 4543Nostoc sp PCC7120 protein	BA000019.2 BAB76247.1	gene=psbD	protein=photosystem II protein D2	protein id=BAB76247.1	location=5441150..5442205
14 1797Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01620.1	gene=psbD2	protein=photosystem II reaction center D2 protein	protein id=BAG01620.1	location=1613898..1614953
14 4112Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03938.1	gene=psbD1	protein=photosystem II reaction center D2 protein	protein id=BAG03938.1	location=complement(3780970..3782025)
15 747Nostoc azollae 0708 protein	CP002059.1 ADI63416.1	gene=Aazo 1059	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ADI63416.1	location=1131903..1132958
15 2645Nostoc azollae 0708 protein	CP002059.1 ADI65314.1	gene=Aazo 3778	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ADI65314.1	location=3834890..3835945
16 1031Acharyochloris marina MBIC11017 protein	CP000828.1 ABW26122.1	gene=psbD	protein=photosystem II D2 protein PsbD	protein id=ABW26122.1	location=1050536..1051591
16 3977Acharyochloris marina MBIC11017 protein	CP000828.1 ABW29065.1	gene=psbD	protein=photosystem II D2 protein PsbD	protein id=ABW29065.1	location=complement(4131304..4132359)
17 3223Nostoc punctiforme PCC73102 protein	CP001037.1 ACC82027.1	gene=Npun R3637	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ACC82027.1	location=complement(4583337..4584392)
17 4110Nostoc punctiforme PCC73102 protein	CP001037.1 ACC82914.1	gene=Npun F4553	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ACC82914.1	location=5658199..5659254
18 8735synechococcus elongatus PCC6301 protein	AP008231.1 BAD79063.1	gene=psbDI	protein=photosystem II reaction center D2 protein	protein id=BAD79063.1	location=complement(969581..970639)
18 24475synechococcus elongatus PCC6301 protein	AP008231.1 BAD80638.1	gene=psbDII	protein=photosystem II reaction center D2 protein	protein id=BAD80638.1	location=2610565..2611623
21 16475synechococcus sp PCC7803 protein	CT971583.1 CAK24073.1	gene=psbD	protein=Photosystem II D2 protein	protein id=CAK24073.1	location=1513257..1514312
21 22395synechococcus sp PCC7803 protein	CT971583.1 CAK24665.1	gene=psbD	protein=Photosystem II D2 protein	protein id=CAK24665.1	location=complement(2064063..2065118)
22 1533Synechococcus sp PCC7002 protein	CP000951.1 ACA99551.1	gene=psbD	protein=photosystem II D2 protein (photosystem q(a) protein)	protein id=ACA99551.1	location=complement(1646526..1647584)
22 2161Synechococcus sp PCC7002 protein	CP000951.1 ACB00180.1	gene=psbD	protein=photosystem II D2 protein	protein id=ACB00180.1	location=2288194..2289252
1 3242Cyanothece sp PCC8801 protein	CP001287.1 ACK67350.1	gene=PCC8801 3382	protein=transposase IS4 family protein	protein id=ACK67350.1	location=complement(3528365..3529384)
1 3688Cyanothece sp PCC8801 protein	CP001287.1 ACK67796.1	gene=PCC8801 3846	protein=transposase IS4 family protein	protein id=ACK67796.1	location=complement(4025990..4027009)
5 261Anabaena variabilis ATCC29413 protein	CP000117.1 ABA19887.1	gene=Ava 0261	protein=transposase, IS4 family	protein id=ABA19887.1	location=334202..335224

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
5 3135Anabaena variabilis ATCC29413 protein	CP000117.1 ABA22761.1	gene=Ava 3153	protein=transposase, IS4 family	protein id=ABA22761.1	location=3914670..3915692
7 1078Synechocystis sp PCC6803 protein	BA000022.2 BAA17657.1	gene=sll1780	protein=transposase	protein id=BAA17657.1	location=complement(1200376..1201392)
7 1576Synechocystis sp PCC6803 protein	BA000022.2 BAA18155.1	gene=sll1255	protein=transposase	protein id=BAA18155.1	location=complement(1729012..1730028)
7 1793Synechocystis sp PCC6803 protein	BA000022.2 BAA18372.1	gene=sll1560	protein=transposase	protein id=BAA18372.1	location=complement(1970952..1971968)
7 1868Synechocystis sp PCC6803 protein	BA000022.2 BAA18447.1	gene=slr1635	protein=transposase	protein id=BAA18447.1	location=2048497..2049513
7 2098Synechocystis sp PCC6803 protein	BA000022.2 BAA10067.1	gene=slr0180	protein=transposase	protein id=BAA10067.1	location=2327013..2328029
7 3020Synechocystis sp PCC6803 protein	BA000022.2 BAA18731.1	gene=sll1474	protein=transposase	protein id=BAA18731.1	location=complement(3400402..3401418)
8 761Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88702.1	gene=glr0761	protein=glr0761	protein id=BAC88702.1	location=813826..814881
8 1195Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89136.1	gene=gll1195	protein=gll1195	protein id=BAC89136.1	location=complement(1271101..1272156)
17 1266Nostoc punctiforme PCC73102 protein	CP001037.1 ACC80069.1	gene=Npun F1359	protein=transposase, IS4 family protein	protein id=ACC80069.1	location=1653292..1654314
17 1374Nostoc punctiforme PCC73102 protein	CP001037.1 ACC80177.1	gene=Npun F1481	protein=transposase, IS4 family protein	protein id=ACC80177.1	location=1818320..1819342
17 2675Nostoc punctiforme PCC73102 protein	CP001037.1 ACC81479.1	gene=Npun F2948	protein=transposase, IS4 family protein	protein id=ACC81479.1	location=3648697..3649719
17 3118Nostoc punctiforme PCC73102 protein	CP001037.1 ACC81922.1	gene=Npun F3516	protein=transposase, IS4 family protein	protein id=ACC81922.1	location=4420430..4421452
17 3218Nostoc punctiforme PCC73102 protein	CP001037.1 ACC82022.1	gene=Npun F3628	protein=transposase, IS4 family protein	protein id=ACC82022.1	location=4572908..4573930
19 761Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88702.1	gene=glr0761	protein=glr0761	protein id=BAC88702.1	location=813826..814881
19 1195Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89136.1	gene=gll1195	protein=gll1195	protein id=BAC89136.1	location=complement(1271101..1272156)
8 872Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88813.1	gene=glr0872	protein=glr0872	protein id=BAC88813.1	location=919913..920968
8 1093Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89034.1	gene=gll1093	protein=gll1093	protein id=BAC89034.1	location=complement(1171434..1172489)
8 2126Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC90067.1	gene=gll2126	protein=gll2126	protein id=BAC90067.1	location=complement(2280085..2281140)
19 872Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88813.1	gene=glr0872	protein=glr0872	protein id=BAC88813.1	location=919913..920968
19 1093Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89034.1	gene=gll1093	protein=gll1093	protein id=BAC89034.1	location=complement(1171434..1172489)
19 2126Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90067.1	gene=gll2126	protein=gll2126	protein id=BAC90067.1	location=complement(2280085..2281140)
8 209Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88150.1	gene=gll0209	protein=gll0209	protein id=BAC88150.1	location=complement(202099..203154)
8 1105Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89046.1	gene=gll1105	protein=gll1105	protein id=BAC89046.1	location=complement(1185264..1186319)
19 209Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88150.1	gene=gll0209	protein=gll0209	protein id=BAC88150.1	location=complement(202099..203154)
19 1105Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89046.1	gene=gll1105	protein=gll1105	protein id=BAC89046.1	location=complement(1185264..1186319)
6 3383Arthrospira platensis NIES-39 protein	AP011615.1 BAI91225.1	gene=NIES39 J01730	protein=putative transposase	protein id=BAI91225.1	location=3451691..3452098
6 3707Arthrospira platensis NIES-39 protein	AP011615.1 BAI91549.1	gene=NIES39 J05030	protein=putative transposase	protein id=BAI91549.1	location=3760401..3760823
6 4464Arthrospira platensis NIES-39 protein	AP011615.1 BAI92306.1	gene=NIES39 L01450	protein=putative transposase	protein id=BAI92306.1	location=complement(4567432..4567854)
8 1611Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89552.1	gene=gll1611	protein=gll1611	protein id=BAC89552.1	location=complement(1728502..1728960)
8 1727Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89668.1	gene=glr1727	protein=glr1727	protein id=BAC89668.1	location=1833160..1833618
8 4074Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC92017.1	gene=glr4076	protein=glr4076	protein id=BAC92017.1	location=4271569..4272027
16 1072Acharyochloris marina MBIC11017 protein	CP000828.1 ABW26163.1	gene=AM1 1124	protein=transposase, putative	protein id=ABW26163.1	location=1099836..1100309
16 1257Acharyochloris marina MBIC11017 protein	CP000828.1 ABW26348.1	gene=AM1 1314	protein=transposase, putative	protein id=ABW26348.1	location=1300852..1301325
16 3820Acharyochloris marina MBIC11017 protein	CP000828.1 ABW28908.1	gene=AM1 3923	protein=conserved hypothetical protein	protein id=ABW28908.1	location=3974665..3975138
19 1611Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89552.1	gene=gll1611	protein=gll1611	protein id=BAC89552.1	location=complement(1728502..1728960)

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
19 1727Gloeobacter violaceus PCC7421 protein	BAA0000452 BAC89668.1	gene=glr1727	protein=glr1727	protein id=BAC89668.1	location=1833160..1833618
19 4074Gloeobacter violaceus PCC7421 protein	BAA0000452 BAC92017.1	gene=glr4076	protein=glr4076	protein id=BAC92017.1	location=4271569..4272027
20 1086Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG50616.1	gene=Tery 1283	protein=hypothetical protein	protein id=ABG50616.1	location=1963943..1964374
20 1992Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG51522.1	gene=Tery 2298	protein=hypothetical protein	protein id=ABG51522.1	location=complement(3567237..3567647)
20 2498Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG52028.1	gene=Tery 2854	protein=hypothetical protein	protein id=ABG52028.1	location=complement(4445965..4446375)
20 2672Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG52203.1	gene=Tery 3055	protein=hypothetical protein	protein id=ABG52203.1	location=complement(4715752..4716162)
16 452Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25543.1	gene=AM1 0490	protein=transposase, putative	protein id=ABW25543.1	location=469494..469901
16 462Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25553.1	gene=AM1 0500	protein=transposase, putative	protein id=ABW25553.1	location=complement(477089..477496)
16 2018Acharyochloris marina MBIC11017 protein	CP000828.1 ABW27109.1	gene=AM1 2094	protein=conserved hypothetical protein	protein id=ABW27109.1	location=2086211..2086618
5 798Anabaena variabilis ATCC29413 protein	CP000117.1 ABA20424.1	gene=Ava 0800	protein=transposase	protein id=ABA20424.1	location=981653..982036
5 1016Anabaena variabilis ATCC29413 protein	CP000117.1 ABA20642.1	gene=Ava 1018	protein=transposase	protein id=ABA20642.1	location=1228979..1229362
5 1362Anabaena variabilis ATCC29413 protein	CP000117.1 ABA20988.1	gene=Ava 1364	protein=transposase	protein id=ABA20988.1	location=complement(1680843..1681226)
5 2629Anabaena variabilis ATCC29413 protein	CP000117.1 ABA22255.1	gene=Ava 2641	protein=transposase	protein id=ABA22255.1	location=3269211..3269594
5 3298Anabaena variabilis ATCC29413 protein	CP000117.1 ABA22924.1	gene=Ava 3317	protein=transposase	protein id=ABA22924.1	location=4141228..4141611
5 3779Anabaena variabilis ATCC29413 protein	CP000117.1 ABA23405.1	gene=Ava 3800	protein=transposase	protein id=ABA23405.1	location=4735420..4735803
7 7845synechocystis sp PCC6803 protein	BAA0000222 BAA17363.1	gene=sll1791	protein=transposase	protein id=BAA17363.1	location=complement(852895..853281)
7 27415synechocystis sp PCC6803 protein	BAA0000222 BAA10711.1	gene=slr0799	protein=transposase	protein id=BAA10711.1	location=3066429..3066752
12 16Nostoc sp PCC7120 protein	BAA0000192 BAB77540.1	gene=all0016	protein=transposase	protein id=BAB77540.1	location=complement(13138..13521)
12 2689Nostoc sp PCC7120 protein	BAA0000192 BAB74392.1	gene=all2693	protein=transposase	protein id=BAB74392.1	location=complement(3284646..3285029)
12 3606Nostoc sp PCC7120 protein	BAA0000192 BAB75309.1	gene=all3610	protein=transposase	protein id=BAB75309.1	location=4361124..4361507
12 4395Nostoc sp PCC7120 protein	BAA0000192 BAB76099.1	gene=all4400	protein=transposase	protein id=BAB76099.1	location=complement(5276131..5276514)
12 4433Nostoc sp PCC7120 protein	BAA0000192 BAB76137.1	gene=alr4438	protein=transposase	protein id=BAB76137.1	location=5322370..5322753
12 4812Nostoc sp PCC7120 protein	BAA0000192 BAB76516.1	gene=all4817	protein=transposase	protein id=BAB76516.1	location=complement(5735861..5736244)
12 5152Nostoc sp PCC7120 protein	BAA0000192 BAB76856.1	gene=alr5157	protein=transposase	protein id=BAB76856.1	location=6156759..6157142
15 3396Nostoc azollae 0708 protein	CP002059.1 ADI66065.1	gene=Aazo 4932	protein=transposase	protein id=ADI66065.1	location=complement(5047600..5047983)
15 3550Nostoc azollae 0708 protein	CP002059.1 ADI66219.1	gene=Aazo 5164	protein=transposase	protein id=ADI66219.1	location=5290637..5291020
15 322Nostoc azollae 0708 protein	CP002059.1 ADI62991.1	gene=Aazo 0453	protein=transposase	protein id=ADI62991.1	location=468637..469020
15 3126Nostoc azollae 0708 protein	CP002059.1 ADI65795.1	gene=Aazo 4525	protein=transposase	protein id=ADI65795.1	location=4617611..4617994
6 4008Arthrospira platensis NIES-39 protein	API11615.1 BAI91850.1	gene=NIES539 K02030	protein=putative transposase	protein id=BAI91850.1	location=4067817..4068176
6 4011Arthrospira platensis NIES-39 protein	API11615.1 BAI91853.1	gene=NIES539 K02060	protein=putative transposase	protein id=BAI91853.1	location=complement(4069017..4069376)
8 16125ynechococcus sp JA-3-3Ab protein	BAA0000452 BAC89553.1	gene=gll1612	protein=gll1612	protein id=BAC89553.1	location=complement(1728993..1729379)
8 17265ynechococcus sp JA-3-3Ab protein	BAA0000452 BAC89667.1	gene=glr1726	protein=glr1726	protein id=BAC89667.1	location=complement(1832741..1833127
8 40735ynechococcus sp JA-3-3Ab protein	BAA0000452 BAC92016.1	gene=glr4075	protein=glr4075	protein id=BAC92016.1	location=4271150..4271536
19 1612Gloeobacter violaceus PCC7421 protein	BAA0000452 BAC89553.1	gene=gll1612	protein=gll1612	protein id=BAC89553.1	location=complement(1728993..1729379)
19 1726Gloeobacter violaceus PCC7421 protein	BAA0000452 BAC89667.1	gene=glr1726	protein=glr1726	protein id=BAC89667.1	location=1832741..1833127
19 4073Gloeobacter violaceus PCC7421 protein	BAA0000452 BAC92016.1	gene=glr4075	protein=glr4075	protein id=BAC92016.1	location=4271150..4271536
6 2123Arthrospira platensis NIES-39 protein	API11615.1 BAI89965.1	gene=NIES539 D05480	protein=putative transposase	protein id=BAI89965.1	location=2128937..2129296
6 3647Arthrospira platensis NIES-39 protein	API11615.1 BAI91489.1	gene=NIES539 J04420	protein=putative transposase	protein id=BAI91489.1	location=complement(3710514..3710873)
20 1085Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG50615.1	gene=Tery 1282	protein=putative transposase	protein id=ABG50615.1	location=1963528..1963887
20 1993Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG51523.1	gene=Tery 2299	protein=putative transposase	protein id=ABG51523.1	location=complement(3567724..3568083)

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
20 2499Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG52029.1	gene=Tery 2855	protein=putative transposase	protein id=ABG52029.1	location=complement(4446452..4446811)
20 2673Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG52204.1	gene=Tery 3056	protein=putative transposase	protein id=ABG52204.1	location=complement(4716239..4716598)
1 1353Cyanothece sp PCC8801 protein	CP001287.1 ACK65460.1	gene=PCC8801 1400	protein=RNA-directed DNA polymerase (Reverse transcriptase)	protein id=ACK65460.1	location=1464222..1466045
1 1354Cyanothece sp PCC8801 protein	CP001287.1 ACK65461.1	gene=PCC8801 1402	protein=RNA-directed DNA polymerase (Reverse transcriptase)	protein id=ACK65461.1	location=1466831..1468654
1 2846Cyanothece sp PCC8801 protein	CP001287.1 ACK66954.1	gene=PCC8801 2962	protein=RNA-directed DNA polymerase (Reverse transcriptase)	protein id=ACK66954.1	location=3059473..3061296
1 2847Cyanothece sp PCC8801 protein	CP001287.1 ACK66955.1	gene=PCC8801 2964	protein=RNA-directed DNA polymerase (Reverse transcriptase)	protein id=ACK66955.1	location=3062082..3063905
6 452Arthrospira platensis NIES-39 protein	AP011615.1 BAI88294.1	gene=NIES39 A04560	protein=reverse transcriptase homolog	protein id=BAI88294.1	location=455495..457261
6 998Arthrospira platensis NIES-39 protein	AP011615.1 BAI88840.1	gene=NIES39 B00830	protein=reverse transcriptase homolog	protein id=BAI88840.1	location=985697..987463
6 1914Arthrospira platensis NIES-39 protein	AP011615.1 BAI89756.1	gene=NIES39 D03370	protein=reverse transcriptase homolog	protein id=BAI89756.1	location=complement(1937168..1938934)
6 4356Arthrospira platensis NIES-39 protein	AP011615.1 BAI92198.1	gene=NIES39 L00370	protein=reverse transcriptase homolog	protein id=BAI92198.1	location=complement(4471823..4473589)
6 5007Arthrospira platensis NIES-39 protein	AP011615.1 BAI92849.1	gene=NIES39 M00110	protein=reverse transcriptase homolog	protein id=BAI92849.1	location=complement(5088414..5090180)
6 6149Arthrospira platensis NIES-39 protein	AP011615.1 BAI93991.1	gene=NIES39 P00110	protein=reverse transcriptase homolog	protein id=BAI93991.1	location=6262865..6264631
9 949Cyanothece sp PCC7424 protein	CP001291.1 ACK69423.1	gene=PCC7424 0968	protein=RNA-directed DNA polymerase (Reverse transcriptase)	protein id=ACK69423.1	location=complement(1073514..1075460)
9 3260Cyanothece sp PCC7424 protein	CP001291.1 ACK71735.1	gene=PCC7424 3336	protein=RNA-directed DNA polymerase (Reverse transcriptase)	protein id=ACK71735.1	location=3733467..3735413
14 585Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00407.1	gene=MAE 05850	protein=putative group II intron/maturase	protein id=BAG00407.1	location=511461..513305
14 1995Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01818.1	gene=MAE 19960	protein=RNA-directed DNA polymerase	protein id=BAG01818.1	location=complement(1788827..1790671)
14 6204Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG06027.1	gene=MAE 62050	protein=probable reverse transcriptase	protein id=BAG06027.1	location=5747312..5749156
20 196Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG49726.1	gene=Tery 0239	protein=RNA-directed DNA polymerase	protein id=ABG49726.1	location=373560..375467
20 2880Trichodesmium erythraeum IMS101 protein	CP000393.1 ABG52411.1	gene=Tery 3305	protein=RNA-directed DNA polymerase	protein id=ABG52411.1	location=5068307..5070214
5 78Anabaena variabilis ATCC29413 protein	CP000117.1 ABA19704.1	gene=Ava 0078	protein=Gas vesicle protein GVPa	protein id=ABA19704.1	location=complement(100886..101101)
5 79Anabaena variabilis ATCC29413 protein	CP000117.1 ABA19705.1	gene=Ava 0079	protein=Gas vesicle protein GVPa	protein id=ABA19705.1	location=complement(101206..101421)
12 2250Nostoc sp PCC7120 protein	BA000019.2 BAB73952.1	gene=gvpB	protein=gas vesicle protein	protein id=BAB73952.1	location=complement(2706132..2706347)
12 2251Nostoc sp PCC7120 protein	BA000019.2 BAB73953.1	gene=gvpA	protein=gas vesicle protein	protein id=BAB73953.1	location=complement(2706452..2706667)
14 3754Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03580.1	gene=gvpAI	protein=gas vesicle protein GvpAI	protein id=BAG03580.1	location=3399385..3399600
14 3755Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03581.1	gene=gvpAII	protein=gas vesicle protein GvpAII	protein id=BAG03581.1	location=3399955..3400170
14 3756Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03582.1	gene=gvpAIII	protein=gas vesicle protein GvpAIII	protein id=BAG03582.1	location=3400500..3400715
5 1438Anabaena variabilis ATCC29413 protein	CP000117.1 ABA21064.1	gene=Ava 1441	protein=transposase, IS4 family	protein id=ABA21064.1	location=1783089..1783529
5 2475Anabaena variabilis ATCC29413 protein	CP000117.1 ABA22101.1	gene=Ava 2486	protein=transposase, IS4 family	protein id=ABA22101.1	location=complement(3082253..3082642)
12 15Nostoc sp PCC7120 protein	BA000019.2 BAB77539.1	gene=all0015	protein=transposase	protein id=BAB77539.1	location=complement(12728..13168)
12 3607Nostoc sp PCC7120 protein	BA000019.2 BAB75310.1	gene=alr3611	protein=transposase	protein id=BAB75310.1	location=4361477..4361917
12 4394Nostoc sp PCC7120 protein	BA000019.2 BAB76098.1	gene=all4399	protein=transposase	protein id=BAB76098.1	location=complement(5275721..5276161)
12 4434Nostoc sp PCC7120 protein	BA000019.2 BAB76138.1	gene=alr4439	protein=transposase	protein id=BAB76138.1	location=5322723..5323163
12 4811Nostoc sp PCC7120 protein	BA000019.2 BAB76515.1	gene=all4816	protein=transposase	protein id=BAB76515.1	location=complement(5735451..5735891)
12 5153Nostoc sp PCC7120 protein	BA000019.2 BAB76857.1	gene=alr5158	protein=transposase	protein id=BAB76857.1	location=6157112..6157552
15 1014Nostoc azollae 0708 protein	CP002059.1 ADI63683.1	gene=Aazo 1465	protein=transposase IS4 family protein	protein id=ADI63683.1	location=complement(1532502..1532957)
15 1308Nostoc azollae 0708 protein	CP002059.1 ADI63977.1	gene=Aazo 1882	protein=transposase IS4 family protein	protein id=ADI63977.1	location=complement(1955990..1956445)
15 1612Nostoc azollae 0708 protein	CP002059.1 ADI64281.1	gene=Aazo 2318	protein=transposase IS4 family protein	protein id=ADI64281.1	location=complement(2406642..2407097)
8 1184Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89125.1	gene=cpcB	protein=phycocyanin beta chain	protein id=BAC89125.1	location=1263334..1263852
8 3216Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC91158.1	gene=cpcB	protein=phycocyanin beta chain	protein id=BAC91158.1	location=3424755..3425273
18 496Synechococcus elongatus PCC6301 protein	AP008231.1 BAD78686.1	gene=cpcB	protein=phycocyanin beta subunit	protein id=BAD78686.1	location=complement(556210..556731)

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
18 501Synechococcus elongatus PCC6301 protein	AP008231.1 BAD78691.1	gene=cpcB	protein=phycocyanin beta subunit	protein id=BAD78691.1	location=complement(559965..560486)
19 1184Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89125.1	gene=cpcB	protein=phycocyanin beta chain	protein id=BAC89125.1	location=1263334..1263852
19 3216Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC91158.1	gene=cpcB	protein=phycocyanin beta chain	protein id=BAC91158.1	location=3424755..3425273
8 11855ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC89126.1	gene=cpcA	protein=phycocyanin alpha chain	protein id=BAC89126.1	location=1263891..1264379
8 32175ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC91159.1	gene=cpcA	protein=phycocyanin alpha chain	protein id=BAC91159.1	location=3425312..3425800
18 4955ynechococcus elongatus PCC6301 protein	AP008231.1 BAD78685.1	gene=cpcA	protein=phycocyanin alpha subunit	protein id=BAD78685.1	location=complement(555667..556158)
18 5005ynechococcus elongatus PCC6301 protein	AP008231.1 BAD78690.1	gene=cpcA	protein=phycocyanin alpha subunit	protein id=BAD78690.1	location=complement(559422..559913)
19 1185Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89126.1	gene=cpcA	protein=phycocyanin alpha chain	protein id=BAC89126.1	location=1263891..1264379
19 3217Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC91159.1	gene=cpcA	protein=phycocyanin alpha chain	protein id=BAC91159.1	location=3425312..3425800
8 1465ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC88087.1	gene=glr0146	protein=glr0146	protein id=BAC88087.1	location=137192..138595
8 3715ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC88312.1	gene=gll0371	protein=gll0371	protein id=BAC88312.1	location=complement(385225..386628)
19 146Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88087.1	gene=glr0146	protein=glr0146	protein id=BAC88087.1	location=137192..138595
19 371Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88312.1	gene=gll0371	protein=gll0371	protein id=BAC88312.1	location=complement(385225..386628)
8 13875ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC89328.1	gene=glr1387	protein=glr1387	protein id=BAC89328.1	location=1491048..1491683
8 24025ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC90343.1	gene=glr2402	protein=MotA/TolQ/ExbB family proton channel protein	protein id=BAC90343.1	location=2561704..2562339
19 1387Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89328.1	gene=glr1387	protein=glr1387	protein id=BAC89328.1	location=1491048..1491683
19 2402Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90343.1	gene=glr2402	protein=MotA/TolQ/ExbB family proton channel protein	protein id=BAC90343.1	location=2561704..2562339
8 13895ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC89330.1	gene=glr1389	protein=glr1389	protein id=BAC89330.1	location=1492134..1492832
8 24045ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC90345.1	gene=glr2404	protein=glr2404	protein id=BAC90345.1	location=2562721..2563491
19 1389Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89330.1	gene=glr1389	protein=glr1389	protein id=BAC89330.1	location=1492134..1492832
19 2404Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90345.1	gene=glr2404	protein=glr2404	protein id=BAC90345.1	location=2562721..2563491
8 21315ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC90072.1	gene=glr2131	protein=glr2131	protein id=BAC90072.1	location=2283239..2283793
8 25635ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC90504.1	gene=gll2563	protein=gll2563	protein id=BAC90504.1	location=complement(2715339..2715977)
19 2131Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90072.1	gene=glr2131	protein=glr2131	protein id=BAC90072.1	location=2283239..2283793
19 2563Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90504.1	gene=gll2563	protein=gll2563	protein id=BAC90504.1	location=complement(2715339..2715977)
8 32675ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC91209.1	gene=glr3268	protein=glr3268	protein id=BAC91209.1	location=3470391..3470822
8 33465ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC91288.1	gene=gll3347	protein=gll3347	protein id=BAC91288.1	location=complement(3533745..3554176)
19 3267Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC91209.1	gene=glr3268	protein=glr3268	protein id=BAC91209.1	location=3470391..3470822
19 3346Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC91288.1	gene=gll3347	protein=gll3347	protein id=BAC91288.1	location=complement(3533745..3554176)
8 1515ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC88092.1	gene=gll0151	protein=gll0151	protein id=BAC88092.1	location=complement(141822..143390)
8 6065ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC88547.1	gene=gll0606	protein=gll0606	protein id=BAC88547.1	location=complement(649344..650912)
8 29845ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC90925.1	gene=gll2984	protein=gll2984	protein id=BAC90925.1	location=complement(3183107..3184675)
8 38345ynechococcus sp JA-3-3Ab protein	BA000045.2 BAC91777.1	gene=gll3836	protein=gll3836	protein id=BAC91777.1	location=complement(4039599..4041167)

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
8 4320Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC92263.1	gene=glr4322	protein=glr4322	protein id=BAC92263.1	location=4553924..4555492
19 151Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88092.1	gene=gll0151	protein=gll0151	protein id=BAC88092.1	location=complement(141822..143390)
19 606Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88547.1	gene=gll0606	protein=gll0606	protein id=BAC88547.1	location=complement(649344..650912)
19 2984Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90925.1	gene=gll2984	protein=gll2984	protein id=BAC90925.1	location=complement(3183107..3184675)
19 3834Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC91777.1	gene=gll3836	protein=gll3836	protein id=BAC91777.1	location=complement(4039599..4041167)
19 4320Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC92263.1	gene=glr4322	protein=glr4322	protein id=BAC92263.1	location=4553924..4555492
8 469Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88410.1	gene=rfbA	protein=glucose-1-phosphate thymidylyltransferase	protein id=BAC88410.1	location=502369..503244
8 1783Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89724.1	gene=gll1783	protein=glucose 1-phosphate thymidyltransferase	protein id=BAC89724.1	location=complement(1897569..1898444)
19 469Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88410.1	gene=rfbA	protein=glucose-1-phosphate thymidylyltransferase	protein id=BAC88410.1	location=502369..503244
19 1783Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89724.1	gene=gll1783	protein=glucose 1-phosphate thymidyltransferase	protein id=BAC89724.1	location=complement(1897569..1898444)
8 518Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88459.1	gene=gll0518	protein=gll0518	protein id=BAC88459.1	location=complement(556855..557763)
8 2561Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC90502.1	gene=gll2561	protein=gll2561	protein id=BAC90502.1	location=complement(2714074..2714982)
8 4190Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC92133.1	gene=glr4192	protein=glr4192	protein id=BAC92133.1	location=4387734..4388642
19 518Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88459.1	gene=gll0518	protein=gll0518	protein id=BAC88459.1	location=complement(556855..557763)
19 2561Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90502.1	gene=gll2561	protein=gll2561	protein id=BAC90502.1	location=complement(2714074..2714982)
19 4190Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC92133.1	gene=glr4192	protein=glr4192	protein id=BAC92133.1	location=4387734..4388642
8 519Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88460.1	gene=gll0519	protein=gll0519	protein id=BAC88460.1	location=complement(557792..558193)
8 2562Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC90503.1	gene=gll2562	protein=gll2562	protein id=BAC90503.1	location=complement(2715011..2715364)
8 4189Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC92132.1	gene=glr4191	protein=glr4191	protein id=BAC92132.1	location=4387304..4387705
19 519Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88460.1	gene=gll0519	protein=gll0519	protein id=BAC88460.1	location=complement(557792..558193)
19 2562Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90503.1	gene=gll2562	protein=gll2562	protein id=BAC90503.1	location=complement(2715011..2715364)
19 4189Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC92132.1	gene=glr4191	protein=glr4191	protein id=BAC92132.1	location=4387304..4387705
8 914Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC88855.1	gene=gll0914	protein=gll0914	protein id=BAC88855.1	location=complement(969401..969739)
8 1495Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89436.1	gene=gll1495	protein=gll1495	protein id=BAC89436.1	location=complement(1608188..1608526)
19 914Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC88855.1	gene=gll0914	protein=gll0914	protein id=BAC88855.1	location=complement(969401..969739)
19 1495Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89436.1	gene=gll1495	protein=gll1495	protein id=BAC89436.1	location=complement(1608188..1608526)
8 1530Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC89471.1	gene=glr1530	protein=pyruvate dehydrogenase E1 component beta	protein id=BAC89471.1	location=1650902..1651885
8 2846Synechococcus sp JA-3-3Ab protein	BA000045.2 BAC90787.1	gene=glr2846	protein=pyruvate dehydrogenase E1 beta-subunit	protein id=BAC90787.1	location=3032535..3033518
19 1530Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC89471.1	gene=glr1530	protein=pyruvate dehydrogenase E1 component beta	protein id=BAC89471.1	location=1650902..1651885
19 2846Gloeobacter violaceus PCC7421 protein	BA000045.2 BAC90787.1	gene=glr2846	protein=pyruvate dehydrogenase E1 beta-subunit	protein id=BAC90787.1	location=3032535..3033518
5 4345Anabaena variabilis ATCC29413 protein	CP000117.1 ABA23972.1	gene=Ava 4374	protein=conserved hypothetical protein	protein id=ABA23972.1	location=5481965..5482207
5 4346Anabaena variabilis ATCC29413 protein	CP000117.1 ABA23973.1	gene=Ava 4375	protein=conserved hypothetical protein	protein id=ABA23973.1	location=5482353..5482589
12 1930Nostoc sp PCC7120 protein	BA000019.2 BAB73632.1	gene=asl1933	protein=asl1933	protein id=BAB73632.1	location=complement(2319911..2320144)
12 1932Nostoc sp PCC7120 protein	BA000019.2 BAB73634.1	gene=asl1935	protein=asl1935	protein id=BAB73634.1	location=complement(2320597..2320830)
7 42Synechocystis sp PCC6803 protein	BA000022.2 BAA16620.1	gene=sll1397	protein=transposase	protein id=BAA16620.1	location=complement(52260..53108)
7 348Synechocystis sp PCC6803 protein	BA000022.2 BAA16927.1	gene=slr1075	protein=transposase	protein id=BAA16927.1	location=379065..379913
7 992Synechocystis sp PCC6803 protein	BA000022.2 BAA17571.1	gene=slr1357	protein=transposase	protein id=BAA17571.1	location=1098323..1099171

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
7 1466Synechocystis sp PCC6803 protein	BA000022.2 BAA18045.1	gene=sll1998	protein=transposase	protein id=BAA18045.1	location=complement(1623723..1624571)
7 2195Synechocystis sp PCC6803 protein	BA000022.2 BAA10164.1	gene=slr0352	protein=transposase	protein id=BAA10164.1	location=2443999..2444847
7 2265Synechocystis sp PCC6803 protein	BA000022.2 BAA10234.1	gene=slr0230	protein=transposase	protein id=BAA10234.1	location=2534106..2534954
7 2772Synechocystis sp PCC6803 protein	BA000022.2 BAA10742.1	gene=slr0704	protein=transposase	protein id=BAA10742.1	location=3097436..3098284
7 3107Synechocystis sp PCC6803 protein	BA000022.2 BAA18818.1	gene=sll0431	protein=transposase	protein id=BAA18818.1	location=complement(3512315..3513163)
14 742Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00564.1	gene=MAE 07420	protein=transposase	protein id=BAG00564.1	location=644876..645724
14 950Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00773.1	gene=MAE 09510	protein=transposase	protein id=BAG00773.1	location=complement(824160..825008)
14 1410Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01233.1	gene=MAE 14110	protein=transposase	protein id=BAG01233.1	location=1255936..1256784
14 1441Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01264.1	gene=MAE 14420	protein=transposase	protein id=BAG01264.1	location=1286000..1286848
14 2041Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01864.1	gene=MAE 20420	protein=transposase	protein id=BAG01864.1	location=complement(1829857..1830705)
14 2400Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02223.1	gene=MAE 24010	protein=transposase	protein id=BAG02223.1	location=complement(2168310..2169158)
14 2464Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02287.1	gene=MAE 24650	protein=transposase	protein id=BAG02287.1	location=2228421..2229269
14 2819Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02642.1	gene=MAE 28200	protein=transposase	protein id=BAG02642.1	location=complement(2567857..2568705)
14 3227Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03052.1	gene=MAE 32300	protein=transposase	protein id=BAG03052.1	location=2946333..2947181
14 3632Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03458.1	gene=MAE 36360	protein=transposase	protein id=BAG03458.1	location=complement(3279361..3280209)
14 3837Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03663.1	gene=MAE 38410	protein=transposase	protein id=BAG03663.1	location=complement(3471789..3472637)
14 3892Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03718.1	gene=MAE 38960	protein=transposase	protein id=BAG03718.1	location=complement(3566138..3566986)
14 4513Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG04339.1	gene=MAE 45170	protein=transposase	protein id=BAG04339.1	location=4154465..4155313
14 5113Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG04937.1	gene=MAE 51150	protein=transposase	protein id=BAG04937.1	location=4697263..4698111
14 5172Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG04996.1	gene=MAE 51740	protein=transposase	protein id=BAG04996.1	location=complement(4746045..4746893)
14 5381Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG05204.1	gene=MAE 53820	protein=transposase	protein id=BAG05204.1	location=complement(4944696..4945544)
14 5503Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG05327.1	gene=MAE 55050	protein=transposase	protein id=BAG05327.1	location=complement(5080677..5081525)
14 5660Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG05484.1	gene=MAE 56620	protein=transposase	protein id=BAG05484.1	location=complement(5221132..5221980)
14 5984Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG05807.1	gene=MAE 59850	protein=transposase	protein id=BAG05807.1	location=5513511..5514359
1 568Cyanothece sp PCC8801 protein	CP001287.1 ACK64675.1	gene=PCC8801 0587	protein=transposase IS200-family protein	protein id=ACK64675.1	location=615742..616155
1 813Cyanothece sp PCC8801 protein	CP001287.1 ACK64920.1	gene=PCC8801 0839	protein=transposase IS200-family protein	protein id=ACK64920.1	location=complement(878858..879271)
1 1294Cyanothece sp PCC8801 protein	CP001287.1 ACK65401.1	gene=PCC8801 1338	protein=transposase IS200-family protein	protein id=ACK65401.1	location=1393277..1393690
9 169Cyanothece sp PCC7424 protein	CP001291.1 ACK68643.1	gene=PCC7424 0173	protein=transposase IS200-family protein	protein id=ACK68643.1	location=191551..191970
9 1400Cyanothece sp PCC7424 protein	CP001291.1 ACK69875.1	gene=PCC7424 1431	protein=transposase IS200-family protein	protein id=ACK69875.1	location=complement(1574976..1575395)
9 2731Cyanothece sp PCC7424 protein	CP001291.1 ACK71206.1	gene=PCC7424 2799	protein=transposase IS200-family protein	protein id=ACK71206.1	location=3111086..3111505
3 134Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC07687.1	gene=tlr0134	protein=tlr0134	protein id=BAC07687.1	location=114353..115579
3 205Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC07758.1	gene=tll0205	protein=tll0205	protein id=BAC07758.1	location=complement(185743..186915)
3 222Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC07775.1	gene=tll0222	protein=tll0222	protein id=BAC07775.1	location=complement(202481..203653)
3 273Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC07826.1	gene=tlr0273	protein=tlr0273	protein id=BAC07826.1	location=257476..258648

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
3 405Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC07958.1	gene=tlr0406	protein=tlr0406	protein id=BAC07958.1	location=403394..404566
3 511Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08064.1	gene=tll0512	protein=tll0512	protein id=BAC08064.1	location=complement(508223..509395)
3 603Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08156.1	gene=tll0604	protein=tll0604	protein id=BAC08156.1	location=complement(611416..612588)
3 685Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08238.1	gene=tll0687	protein=tll0687	protein id=BAC08238.1	location=complement(706573..707799)
3 1125Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08679.1	gene=tlr1127	protein=tlr1127	protein id=BAC08679.1	location=1159903..1161090
3 1150Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08704.1	gene=tll1152	protein=tll1152	protein id=BAC08704.1	location=complement(1183861..1185033)
3 1372Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08926.1	gene=tll1374	protein=tll1374	protein id=BAC08926.1	location=complement(1441028..1442200)
3 2177Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC09731.1	gene=tlr2179	protein=tlr2179	protein id=BAC09731.1	location=2263217..2264389
3 2180Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC09734.1	gene=tlr2182	protein=tlr2182	protein id=BAC09734.1	location=2266280..2267452
3 2197Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC09751.1	gene=tlr2199	protein=tlr2199	protein id=BAC09751.1	location=2283094..2284266
3 2374Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC09928.1	gene=tlr2376	protein=tlr2376	protein id=BAC09928.1	location=2486239..2487411
14 1575Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01398.1	gene=MAE 15760	protein=transposase	protein id=BAG01398.1	location=1415997..1417040
14 2940Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02765.1	gene=MAE 29430	protein=transposase	protein id=BAG02765.1	location=2681481..2682524
6 3611Arthrospira platensis NIES-39 protein	AP011615.1 BAI91453.1	gene=NIES39 J04060	protein=transposase	protein id=BAI91453.1	location=complement(3681424..3681828)
6 5325Arthrospira platensis NIES-39 protein	AP011615.1 BAI93167.1	gene=NIES39 N00500	protein=transposase	protein id=BAI93167.1	location=5473561..5473965
9 1437Cyanothece sp PCC7424 protein	CP001291.1 ACK69912.1	gene=PCC7424 1470	protein=transposase IS200-family protein	protein id=ACK69912.1	location=complement(1620393..1620800)
9 5155Cyanothece sp PCC7424 protein	CP001291.1 ACK73630.1	gene=PCC7424 5282	protein=transposase IS200-family protein	protein id=ACK73630.1	location=complement(5859213..5859620)
6 1071Arthrospira platensis NIES-39 protein	AP011615.1 BAI88913.1	gene=NIES39 C00430	protein=hypothetical protein	protein id=BAI88913.1	location=complement(1065361..1066566)
6 5156Arthrospira platensis NIES-39 protein	AP011615.1 BAI92998.1	gene=NIES39 M01610	protein=hypothetical protein	protein id=BAI92998.1	location=5245558..5246763
14 1122Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00945.1	gene=MAE 11230	protein=hypothetical protein	protein id=BAG00945.1	location=complement(974280..975422)
14 4916Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG04740.1	gene=MAE 49180	protein=hypothetical protein	protein id=BAG04740.1	location=4504901..4505998
1 623Cyanothece sp PCC8801 protein	CP001287.1 ACK64730.1	gene=PCC8801 0643	protein=transposase	protein id=ACK64730.1	location=complement(670640..671638)
1 1928Cyanothece sp PCC8801 protein	CP001287.1 ACK66035.1	gene=PCC8801 1997	protein=transposase	protein id=ACK66035.1	location=2072306..2073304
14 1070Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00893.1	gene=MAE 10710	protein=transposase	protein id=BAG00893.1	location=complement(918742..919740)
14 1388Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01211.1	gene=MAE 13890	protein=transposase	protein id=BAG01211.1	location=1239551..1240408
14 1979Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01802.1	gene=MAE 19800	protein=transposase	protein id=BAG01802.1	location=1775638..1776636
14 2026Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01849.1	gene=MAE 20270	protein=transposase and inactivated derivatives	protein id=BAG01849.1	location=complement(1817962..1818960)
14 2223Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02046.1	gene=MAE 22240	protein=transposase	protein id=BAG02046.1	location=1994117..1994884
14 2341Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02164.1	gene=MAE 23420	protein=transposase	protein id=BAG02164.1	location=2108561..2109559
14 2749Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02572.1	gene=MAE 27500	protein=transposase	protein id=BAG02572.1	location=complement(2481279..2482277)
14 2882Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02707.1	gene=MAE 28850	protein=transposase	protein id=BAG02707.1	location=complement(2629909..2630907)
14 3636Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03462.1	gene=MAE 36400	protein=transposase	protein id=BAG03462.1	location=3282649..3283647
14 4403Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG04229.1	gene=MAE 44070	protein=transposase	protein id=BAG04229.1	location=complement(4060904..4061902)
14 6154Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG05977.1	gene=MAE 61550	protein=transposase	protein id=BAG05977.1	location=5699143..5700141
9 1181Cyanothece sp PCC7424 protein	CP001291.1 ACK69656.1	gene=PCC7424 1208	protein=Insertion element protein	protein id=ACK69656.1	location=join(1322136..1322465,1322467..1322829)
9 2458Cyanothece sp PCC7424 protein	CP001291.1 ACK70933.1	gene=PCC7424 2516	protein=Insertion element protein	protein id=ACK70933.1	location=join(2799135..2799464,2799466..2799828)
9 5076Cyanothece sp PCC7424 protein	CP001291.1 ACK73551.1	gene=PCC7424 5203	protein=Insertion element protein	protein id=ACK73551.1	location=complement(join(5775088..5775450,5775452..5775781))
16 695Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25786.1	gene=AM1 0742	protein=IS1 transposase	protein id=ABW25786.1	location=720392..721081
16 3378Acharyochloris marina MBIC11017 protein	CP000828.1 ABW28467.1	gene=AM1 3475	protein=IS1 transposase	protein id=ABW28467.1	location=complement(3510465..3511154)

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
16 5827Acharyochloris marina MBIC11017 protein	CP000828.1 ABW30914.1	gene=AM1 5978	protein=IS1 transposase	protein id=ABW30914.1	location=6055468..6056157
1 1064Cyanothece sp PCC8801 protein	CP001287.1 ACK65171.1	gene=PCC8801 1098	protein=transposase, IS605 OrfB family	protein id=ACK65171.1	location=complement(1157486..1158697)
1 4021Cyanothece sp PCC8801 protein	CP001287.1 ACK68129.1	gene=PCC8801 4201	protein=transposase, IS605 OrfB family	protein id=ACK68129.1	location=4404944..4406155
1 4249Cyanothece sp PCC8801 protein	CP001287.1 ACK68357.1	gene=PCC8801 4435	protein=transposase, IS605 OrfB family	protein id=ACK68357.1	location=complement(4663329..4664540)
3 257Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC07810.1	gene=tlr0257	protein=tlr0257	protein id=BAC07810.1	location=241585..242763
3 400Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC07953.1	gene=tlr0401	protein=tlr0401	protein id=BAC07953.1	location=397671..398849
3 1378Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC08932.1	gene=tlr1380	protein=tlr1380	protein id=BAC08932.1	location=1447751..1448929
3 2236Thermosynechococcus elongatus BP-1 protein	BA000039.2 BAC09790.1	gene=tlr2238	protein=tlr2238	protein id=BAC09790.1	location=2320881..2322059
16 218Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25309.1	gene=AM1 0223	protein=transposase	protein id=ABW25309.1	location=complement(219734..221044)
16 496Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25587.1	gene=AM1 0535	protein=transposase	protein id=ABW25587.1	location=515940..517250
16 637Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25728.1	gene=AM1 0681	protein=transposase	protein id=ABW25728.1	location=666645..667955
16 916Acharyochloris marina MBIC11017 protein	CP000828.1 ABW26007.1	gene=AM1 0965	protein=transposase	protein id=ABW26007.1	location=940229..941539
16 1073Acharyochloris marina MBIC11017 protein	CP000828.1 ABW26164.1	gene=AM1 1125	protein=transposase	protein id=ABW26164.1	location=1100397..1101707
16 1854Acharyochloris marina MBIC11017 protein	CP000828.1 ABW26945.1	gene=AM1 1927	protein=transposase	protein id=ABW26945.1	location=complement(1922983..1924293)
16 2315Acharyochloris marina MBIC11017 protein	CP000828.1 ABW27406.1	gene=AM1 2396	protein=transposase	protein id=ABW27406.1	location=complement(2402873..2404183)
16 2506Acharyochloris marina MBIC11017 protein	CP000828.1 ABW27597.1	gene=AM1 2590	protein=transposase	protein id=ABW27597.1	location=2629004..2630314
16 2509Acharyochloris marina MBIC11017 protein	CP000828.1 ABW27600.1	gene=AM1 2593	protein=transposase	protein id=ABW27600.1	location=complement(2631483..2632793)
16 3530Acharyochloris marina MBIC11017 protein	CP000828.1 ABW28619.1	gene=AM1 3629	protein=transposase	protein id=ABW28619.1	location=complement(3687534..3688844)
16 3602Acharyochloris marina MBIC11017 protein	CP000828.1 ABW28691.1	gene=AM1 3701	protein=transposase	protein id=ABW28691.1	location=complement(3758553..3759863)
16 3821Acharyochloris marina MBIC11017 protein	CP000828.1 ABW28909.1	gene=AM1 3924	protein=transposase	protein id=ABW28909.1	location=3975226..3976536
16 4123Acharyochloris marina MBIC11017 protein	CP000828.1 ABW29211.1	gene=AM1 4231	protein=transposase	protein id=ABW29211.1	location=4250628..4251938
16 5052Acharyochloris marina MBIC11017 protein	CP000828.1 ABW30139.1	gene=AM1 5177	protein=transposase	protein id=ABW30139.1	location=complement(5241022..5242332)
16 5160Acharyochloris marina MBIC11017 protein	CP000828.1 ABW30247.1	gene=AM1 5286	protein=transposase	protein id=ABW30247.1	location=5352546..5353856
16 5539Acharyochloris marina MBIC11017 protein	CP000828.1 ABW30626.1	gene=AM1 5679	protein=transposase	protein id=ABW30626.1	location=complement(5743964..5745274)
16 6106Acharyochloris marina MBIC11017 protein	CP000828.1 ABW31193.1	gene=AM1 6261	protein=transposase	protein id=ABW31193.1	location=complement(6336950..6338260)
17 2423Nostoc punctiforme PCC73102 protein	CP001037.1 ACC81227.1	gene=Npun F2690	protein=Transposase-like protein	protein id=ACC81227.1	location=3341372..3342682
17 2839Nostoc punctiforme PCC73102 protein	CP001037.1 ACC81643.1	gene=Npun F3186	protein=Transposase-like protein	protein id=ACC81643.1	location=3966579..3967889
17 2974Nostoc punctiforme PCC73102 protein	CP001037.1 ACC81778.1	gene=Npun F3348	protein=Transposase-like protein	protein id=ACC81778.1	location=4168700..4170010
17 4991Nostoc punctiforme PCC73102 protein	CP001037.1 ACC83796.1	gene=Npun F5490	protein=Transposase-like protein	protein id=ACC83796.1	location=6779978..6781288
1 1753Cyanothece sp PCC8801 protein	CP001287.1 ACK65860.1	gene=PCC8801 1816	protein=transposase IS204/IS1001/IS1096/IS1165 family protein	protein id=ACK65860.1	location=complement(1898016..1899242)
1 3366Cyanothece sp PCC8801 protein	CP001287.1 ACK67474.1	gene=PCC8801 3509	protein=transposase IS204/IS1001/IS1096/IS1165 family protein	protein id=ACK67474.1	location=3661150..3662376
1 3865Cyanothece sp PCC8801 protein	CP001287.1 ACK67973.1	gene=PCC8801 4034	protein=transposase IS204/IS1001/IS1096/IS1165 family protein	protein id=ACK67973.1	location=complement(4225337..4226563)
14 2469Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02292.1	gene=MAE 24700	protein=transposase	protein id=BAG02292.1	location=complement(2233792..2234811)
14 2696Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02519.1	gene=MAE 26970	protein=transposase	protein id=BAG02519.1	location=complement(2436880..2438094)

strain	accession number	gene annotation	gene description	protein accession number	location in the genome
6 1824Arthrospira platensis NIES-39 protein	AP011615.1 BAI89666.1	gene=NIES39 D02460	protein=putative transposase	protein id=BAI89666.1	location=complement(1848140..1848499)
6 4845Arthrospira platensis NIES-39 protein	AP011615.1 BAI92687.1	gene=NIES39 L05300	protein=putative transposase	protein id=BAI92687.1	location=4909208..4909567
16 451Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25542.1	gene=AM1 0489	protein=transposase, putative	protein id=ABW25542.1	location=469018..469401
16 463Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25554.1	gene=AM1 0501	protein=transposase, putative	protein id=ABW25554.1	location=complement(477589..477972)
16 2017Acharyochloris marina MBIC11017 protein	CP000828.1 ABW27108.1	gene=AM1 2093	protein=transposase, putative	protein id=ABW27108.1	location=2085736..2086095
14 512Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00334.1	gene=MAE 05120	protein=transposase	protein id=BAG00334.1	location=456375..457418
14 748Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00570.1	gene=MAE 07480	protein=transposase	protein id=BAG00570.1	location=649201..650244
14 1030Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00853.1	gene=MAE 10310	protein=transposase	protein id=BAG00853.1	location=complement(884964..886007)
14 4798Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG04622.1	gene=MAE 48000	protein=transposase	protein id=BAG04622.1	location=4401177..4402220
16 124Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25215.1	gene=AM1 0129	protein=hypothetical protein	protein id=ABW25215.1	location=complement(128563..129672)
16 155Acharyochloris marina MBIC11017 protein	CP000828.1 ABW25246.1	gene=AM1 0160	protein=hypothetical protein	protein id=ABW25246.1	location=158975..160084
16 1273Acharyochloris marina MBIC11017 protein	CP000828.1 ABW26364.1	gene=AM1 1330	protein=hypothetical protein	protein id=ABW26364.1	location=1313003..1314112
16 4169Acharyochloris marina MBIC11017 protein	CP000828.1 ABW29257.1	gene=AM1 4277	protein=hypothetical protein	protein id=ABW29257.1	location=complement(4296768..4297877)
16 4971Acharyochloris marina MBIC11017 protein	CP000828.1 ABW30058.1	gene=AM1 5094	protein=hypothetical protein	protein id=ABW30058.1	location=5147289..5148398
16 5777Acharyochloris marina MBIC11017 protein	CP000828.1 ABW30864.1	gene=AM1 5926	protein=hypothetical protein	protein id=ABW30864.1	location=5992224..5993333
16 6105Acharyochloris marina MBIC11017 protein	CP000828.1 ABW31192.1	gene=AM1 6260	protein=hypothetical protein	protein id=ABW31192.1	location=6335818..6336927
14 165Microcystis aeruginosa NIES-843 protein	AP009552.1 BAF99986.1	gene=MAE 01650	protein=hypothetical protein	protein id=BAF99986.1	location=133753..134526
14 3306Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03133.1	gene=MAE 33110	protein=hypothetical protein	protein id=BAG03133.1	location=complement(3016288..3017061)
14 3422Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03248.1	gene=MAE 34260	protein=hypothetical protein	protein id=BAG03248.1	location=complement(3114659..3115432)
14 4305Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG04131.1	gene=MAE 43090	protein=hypothetical protein	protein id=BAG04131.1	location=3971211..3971984
17 1920Nostoc punctiforme PCC73102 protein	CP001037.1 ACC80724.1	gene=Npun F2107	protein=conserved hypothetical protein	protein id=ACC80724.1	location=2576706..2577554
17 4319Nostoc punctiforme PCC73102 protein	CP001037.1 ACC83123.1	gene=Npun R4774	protein=conserved hypothetical protein	protein id=ACC83123.1	location=complement(5915257..5916105)
9 3766Cyanothece sp PCC7424 protein	CP001291.1 ACK72241.1	gene=PCC7424 3862	protein=hypothetical protein	protein id=ACK72241.1	location=complement(join(4297108..4297866,4297868..4297930))
9 3802Cyanothece sp PCC7424 protein	CP001291.1 ACK72277.1	gene=PCC7424 3899	protein=transposase IS702 family protein	protein id=ACK72277.1	location=4332844..4333587
14 484Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00306.1	gene=MAE 04840	protein=transposase	protein id=BAG00306.1	location=431565..432356
14 783Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00605.1	gene=MAE 07830	protein=transposase	protein id=BAG00605.1	location=679549..680406
14 895Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00718.1	gene=MAE 08960	protein=transposase	protein id=BAG00718.1	location=778709..779566
14 928Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00751.1	gene=MAE 09290	protein=transposase	protein id=BAG00751.1	location=803601..804458
14 990Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00813.1	gene=MAE 09910	protein=transposase	protein id=BAG00813.1	location=complement(859757..860614)
14 1086Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG00909.1	gene=MAE 10870	protein=transposase	protein id=BAG00909.1	location=complement(933528..934385)
14 2071Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG01894.1	gene=MAE 20720	protein=transposase	protein id=BAG01894.1	location=1848286..1849143
14 2319Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02142.1	gene=MAE 23200	protein=transposase	protein id=BAG02142.1	location=2084969..2085826
14 2606Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02429.1	gene=MAE 26070	protein=transposase	protein id=BAG02429.1	location=complement(2356803..2357660)
14 2710Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG02533.1	gene=MAE 27110	protein=transposase	protein id=BAG02533.1	location=complement(2448484..2449323)
14 3325Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03152.1	gene=MAE 33300	protein=transposase	protein id=BAG03152.1	location=complement(3028164..3029021)
14 3360Microcystis aeruginosa NIES-843 protein	AP009552.1 BAG03187.1	gene=MAE 33650	protein=transposase	protein id=BAG03187.1	location=3065824..3066591

strain			accession number	gene annotation	gene description	protein accession number	location in the genome
14 3439Microcystis	aeruginosa	NIES-843	AP009552.1 BAG03265.1	gene=MAE 34430	protein=transposase	protein id=BAG03265.1	location=complement(3127634..3128401)
protein							
14 3839Microcystis	aeruginosa	NIES-843	AP009552.1 BAG03665.1	gene=MAE 38430	protein=transposase	protein id=BAG03665.1	location=3473455..3474312
protein							
14 3927Microcystis	aeruginosa	NIES-843	AP009552.1 BAG03753.1	gene=MAE 39310	protein=transposase	protein id=BAG03753.1	location=3616333..3617190
protein							
14 3986Microcystis	aeruginosa	NIES-843	AP009552.1 BAG03812.1	gene=MAE 39900	protein=transposase	protein id=BAG03812.1	location=3663197..3664054
protein							
14 4266Microcystis	aeruginosa	NIES-843	AP009552.1 BAG04092.1	gene=MAE 42700	protein=transposase	protein id=BAG04092.1	location=complement(3938799..3939656)
protein							
14 4843Microcystis	aeruginosa	NIES-843	AP009552.1 BAG04667.1	gene=MAE 48450	protein=transposase	protein id=BAG04667.1	location=4447164..4448021
protein							
14 5305Microcystis	aeruginosa	NIES-843	AP009552.1 BAG05128.1	gene=MAE 53060	protein=transposase	protein id=BAG05128.1	location=complement(4868782..4869639)
protein							
14 5415Microcystis	aeruginosa	NIES-843	AP009552.1 BAG05239.1	gene=MAE 54170	protein=transposase	protein id=BAG05239.1	location=complement(4978973..4979830)
protein							
14 5699Microcystis	aeruginosa	NIES-843	AP009552.1 BAG05523.1	gene=MAE 57010	protein=transposase	protein id=BAG05523.1	location=5261959..5262816
protein							
14 6145Microcystis	aeruginosa	NIES-843	AP009552.1 BAG05968.1	gene=MAE 61460	protein=transposase	protein id=BAG05968.1	location=complement(5692792..5693649)
protein							
14 6233Microcystis	aeruginosa	NIES-843	AP009552.1 BAG06056.1	gene=MAE 62340	protein=transposase	protein id=BAG06056.1	location=complement(5773659..5774516)
protein							
14 6284Microcystis	aeruginosa	NIES-843	AP009552.1 BAG06108.1	gene=MAE 62860	protein=transposase	protein id=BAG06108.1	location=5820820..5821677
protein							
14 1381Microcystis	aeruginosa	NIES-843	AP009552.1 BAG01204.1	gene=MAE 13820	protein=transposase	protein id=BAG01204.1	location=1236257..1236883
protein							
14 2091Microcystis	aeruginosa	NIES-843	AP009552.1 BAG01914.1	gene=MAE 20920	protein=transposase	protein id=BAG01914.1	location=complement(1867658..1868284)
protein							
16 1185Acharyochloris marina MBIC11017			CP000828.1 ABW26276.1	gene=AM1 1238	protein=transposase, putative	protein id=ABW26276.1	location=complement(1225141..1225770)
protein							
16 4321Acharyochloris marina MBIC11017			CP000828.1 ABW29409.1	gene=AM1 4432	protein=transposase, putative	protein id=ABW29409.1	location=complement(4460718..4461347)
protein							
14 145Microcystis	aeruginosa	NIES-843	AP009552.1 BAF99966.1	gene=MAE 01450	protein=transposase	protein id=BAF99966.1	location=120122..120571
protein							
14 2092Microcystis	aeruginosa	NIES-843	AP009552.1 BAG01915.1	gene=MAE 20930	protein=transposase	protein id=BAG01915.1	location=complement(1868347..1868796)
protein							
16 716Acharyochloris marina MBIC11017			CP000828.1 ABW25807.1	gene=AM1 0763	protein=transposase, putative	protein id=ABW25807.1	location=complement(741568..742089)
protein							
16 1481Acharyochloris marina MBIC11017			CP000828.1 ABW26572.1	gene=AM1 1547	protein=transposase, putative	protein id=ABW26572.1	location=1528937..1529458
protein							
16 1545Acharyochloris marina MBIC11017			CP000828.1 ABW26636.1	gene=AM1 1612	protein=transposase, putative	protein id=ABW26636.1	location=1584889..1585410
protein							
12 3982Nostoc sp PCC7120 protein			BA000019.2 BAB75685.1	gene=all3986	protein=transposase	protein id=BAB75685.1	location=complement(4801437..4802645)
12 4100Nostoc sp PCC7120 protein			BA000019.2 BAB75803.1	gene=alr4104	protein=transposase	protein id=BAB75803.1	location=4944752..4945960
14 530Microcystis	aeruginosa	NIES-843	AP009552.1 BAG00352.1	gene=MAE 05300	protein=transposase	protein id=BAG00352.1	location=471056..472243
protein							
14 4484Microcystis	aeruginosa	NIES-843	AP009552.1 BAG04310.1	gene=MAE 44880	protein=transposase	protein id=BAG04310.1	location=complement(4126673..4127860)
protein							
1 364Cyanothece sp PCC8801 protein			CP001287.1 ACK64471.1	gene=PCC8801 0373	protein=putative IS891/IS1136/IS1341 family transposase	protein id=ACK64471.1	location=371253..372587
1 3436Cyanothece sp PCC8801 protein			CP001287.1 ACK67544.1	gene=PCC8801 3581	protein=putative IS891/IS1136/IS1341 family transposase	protein id=ACK67544.1	location=complement(3730986..3732320)
9 804Cyanothece sp PCC7424 protein			CP001291.1 ACK69278.1	gene=PCC7424 0822	protein=putative transposase IS605 family	protein id=ACK69278.1	location=complement(910819..912054)
9 3156Cyanothece sp PCC7424 protein			CP001291.1 ACK71631.1	gene=PCC7424 3231	protein=putative transposase IS605 family	protein id=ACK71631.1	location=3615830..3617065

Additional File 1.2 - 16S rRNA gene copy data including data from the rrndb-database

Table with information on 16S rRNA copy numbers including data received from the rrnDB database [161] marked (*).

Tab. A.1. Taxon Set B

species	# of 16S rRNA copies	cell types
Acharyochloris marina MBIC11017	2	1
Anabaena variabilis ATCC 29413	4	3
Arthrospira platensis NIES 39	2	1
Cyanothece sp. ATCC 51142	2	1
Cyanothece sp. PCC 7424	3	1
Cyanothece sp. PCC 7425	2	1
Cyanothece sp. PCC 7822	3	1
Cyanothece sp. PCC 8801	2	1
Cyanothece sp. PCC 8802	2	1
Gloeobacter violaceus PCC 7421	1	1
Microcystis aeruginosa NIES-843	2	1
Nostoc azollae 0708	4	3
Nostoc punctiforme PCC 73102	4	3
Nostoc sp. PCC 7120	4	3
Prochlorococcus marinus str. AS9601	1	1
Prochlorococcus marinus str. MIT 9211	1	1
Prochlorococcus marinus str. MIT 9215	1	1
Prochlorococcus marinus str. MIT 9301	1	1
Prochlorococcus marinus str. MIT 9303	2	1
Prochlorococcus marinus str. MIT 9312	1	1
Prochlorococcus marinus str. MIT 9313	2	1
Prochlorococcus marinus str. MIT 9515	1	1
Prochlorococcus marinus str. NATL1A	1	1
Prochlorococcus marinus str. NATL2A	1	1
P. marinus subsp. marinus str. CCMP1375	1	1
P. marinus subsp. pastoris str. CCMP1986	1	1
Synechococcus elongatus PCC 6301	2	1
Synechococcus elongatus PCC 7942	2	1
Synechococcus sp. CC9311	2	1
Synechococcus sp. CC9605	2	1
Synechococcus sp. CC9902	2	1
Synechococcus sp. JA-2-3B'a(2-13)	2	1
Synechococcus sp. JA-3-3Ab	2	1
Synechococcus sp. PCC 7002	2	1
Synechococcus sp. RCC307	1	1
Synechococcus sp. WH 7803	2	1
Synechococcus sp. WH 8102	2	1
Synechocystis sp. PCC 6803	2	1
Thermosynechococcus elongatus BP-1	1	1
Trichodesmium erythraeum IMS101	2	1
cyanobacterium UCYN-A	2	1
Oscillatoria sp PCC 6506[1]	1	1
Aphanizomenon flos-aquae PCC 7905[1]	5	3
Anabaenopsis PCC 9215[1]	4	3
Anabaenopsis PCC 9216[1]	4	3
Anabaena flos-aquae PCC 9302[1]	5	3
Anabaena flos-aquae PCC 9332[1]	5	3
Anabaena flos-aquae PCC 9349[1]	5	3
Nodularia PCC 9350[1]	4	3
Anabaenopsis elenkinii PCC 9420[1]	4	3
Cyanospira rippkae PCC 9501[1]	4	3
Cyanospira capsulata PCC 9502[1]	4	3
Anabaenopsis PCC 9608[1]	4	3
Correlation coeffizient of # of copy and sections	r	0.9086627
	p-value	$< 2.2e-16$

Additional File 1.3 - Distribution of 16S rRNA copy numbers using additional data from rrndb

Boxplot representations of the 16S rRNA gene copy number distribution across the previously defined morphological groups. Additional data on 16S rRNA copy numbers were received from the rrndb-database [161]. Spearman's rank correlation coefficient (ρ) and Pearson's correlation coefficient (R) are displayed above the graph. A strong correlation of 16S rRNA gene copies to terminally differentiated cyanobacteria is supported.

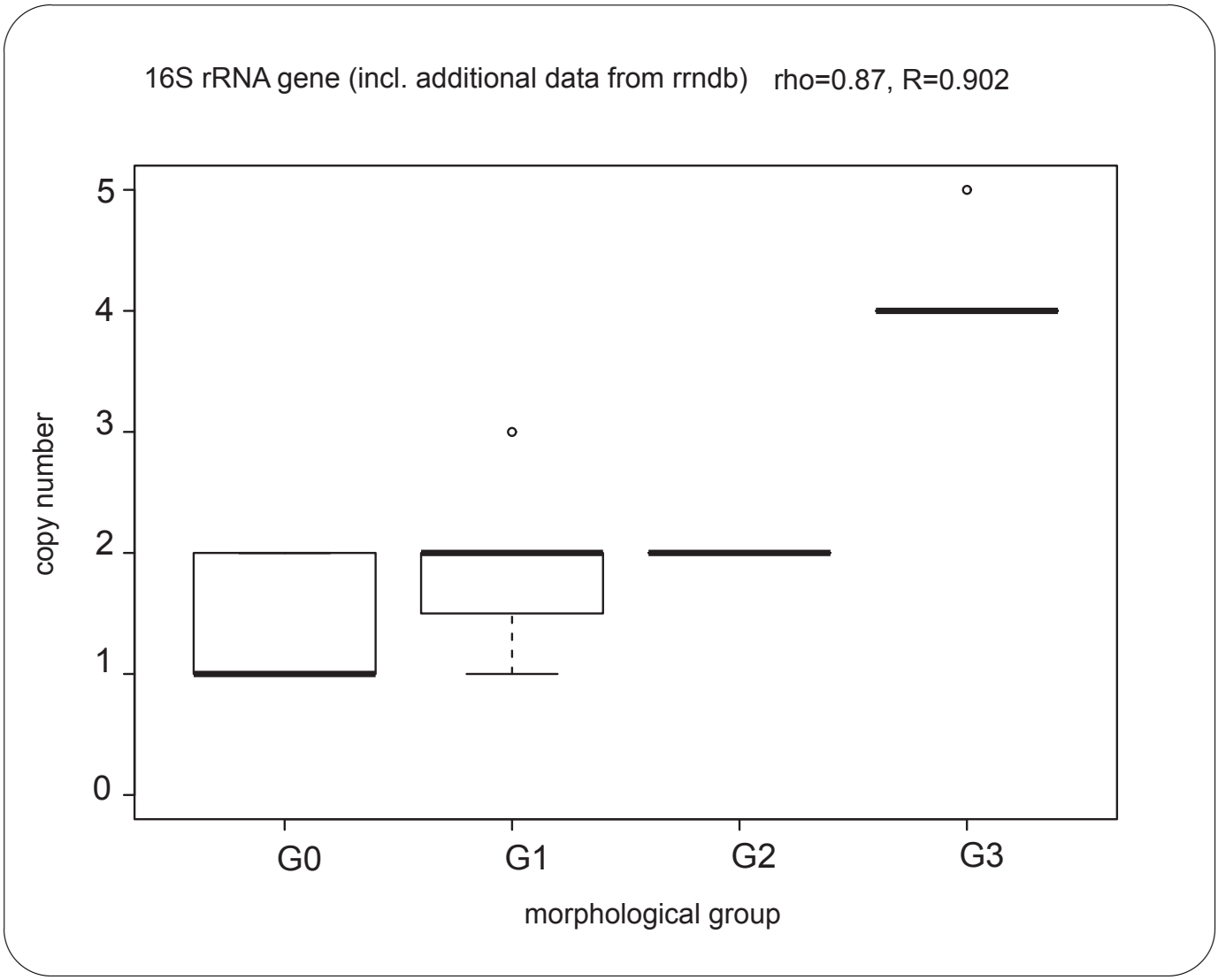


Fig. A.4. Distribution of 16S rRNA copy numbers using additional data from rrndb

Additional File 1.4 - Distribution of mean distances within species from bootstrap samples for the different eubacterial phyla

The distribution of mean distances of the bootstrap samples presented as a histogram. The 95% confidence intervals between cyanobacteria and Chloroflexi, Spirochaetes and Bacteroidetes do not overlap. Cyanobacterial 16S rRNA gene sequence variation within species is significantly lower.

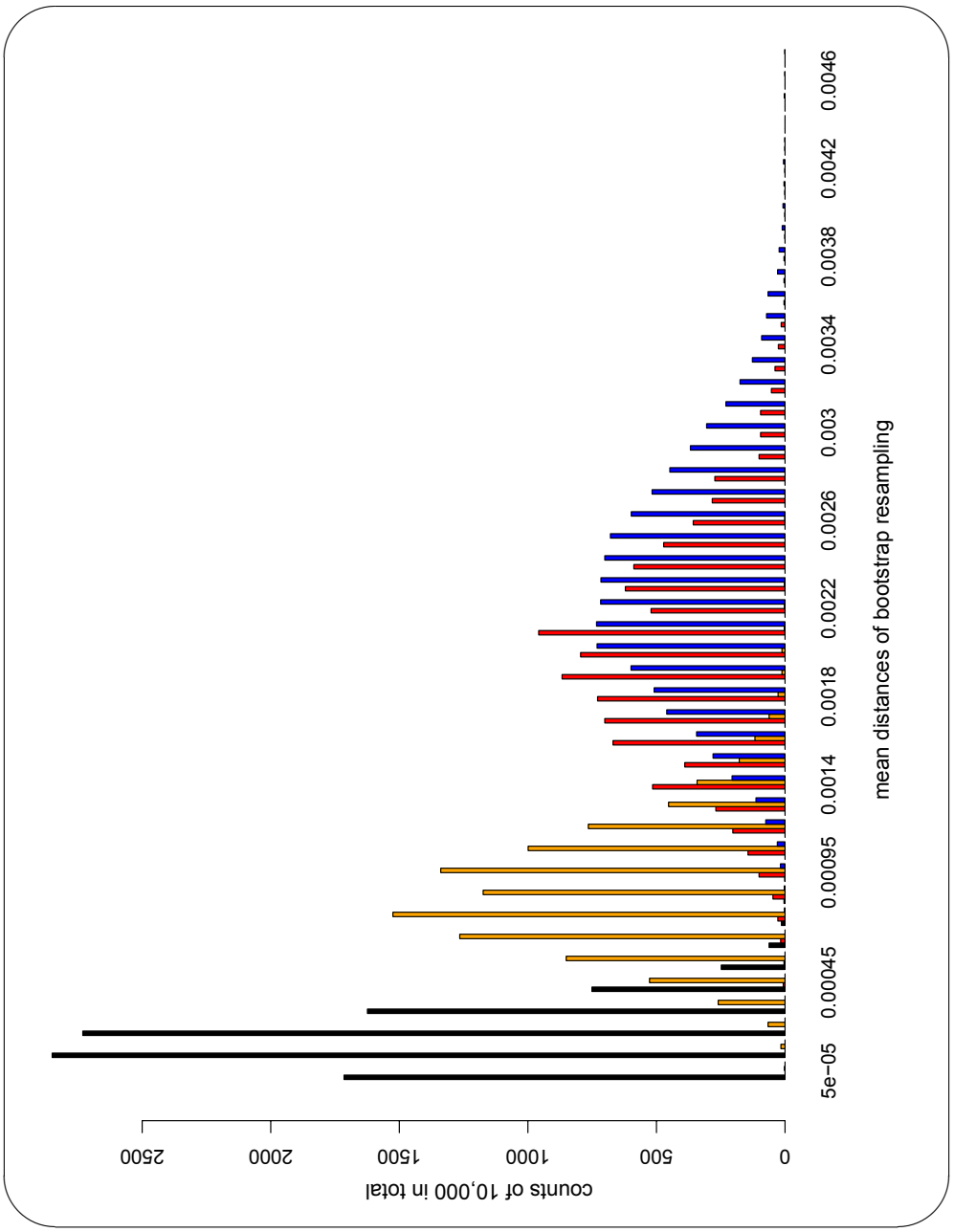


Fig. A.5. Distribution of mean distances within species from bootstrap samples for the different eubacterial phyla

Additional File 1.5 - Distribution of mean distances between species from bootstrap samples for the different eubacterial phyla

The distribution of mean distances of the bootstrap samples presented as a histogram. The 95% confidence intervals between cyanobacteria and the other eubacterial phyla do not overlap. Cyanobacterial 16S rRNA gene sequence variation between species are significantly lower.

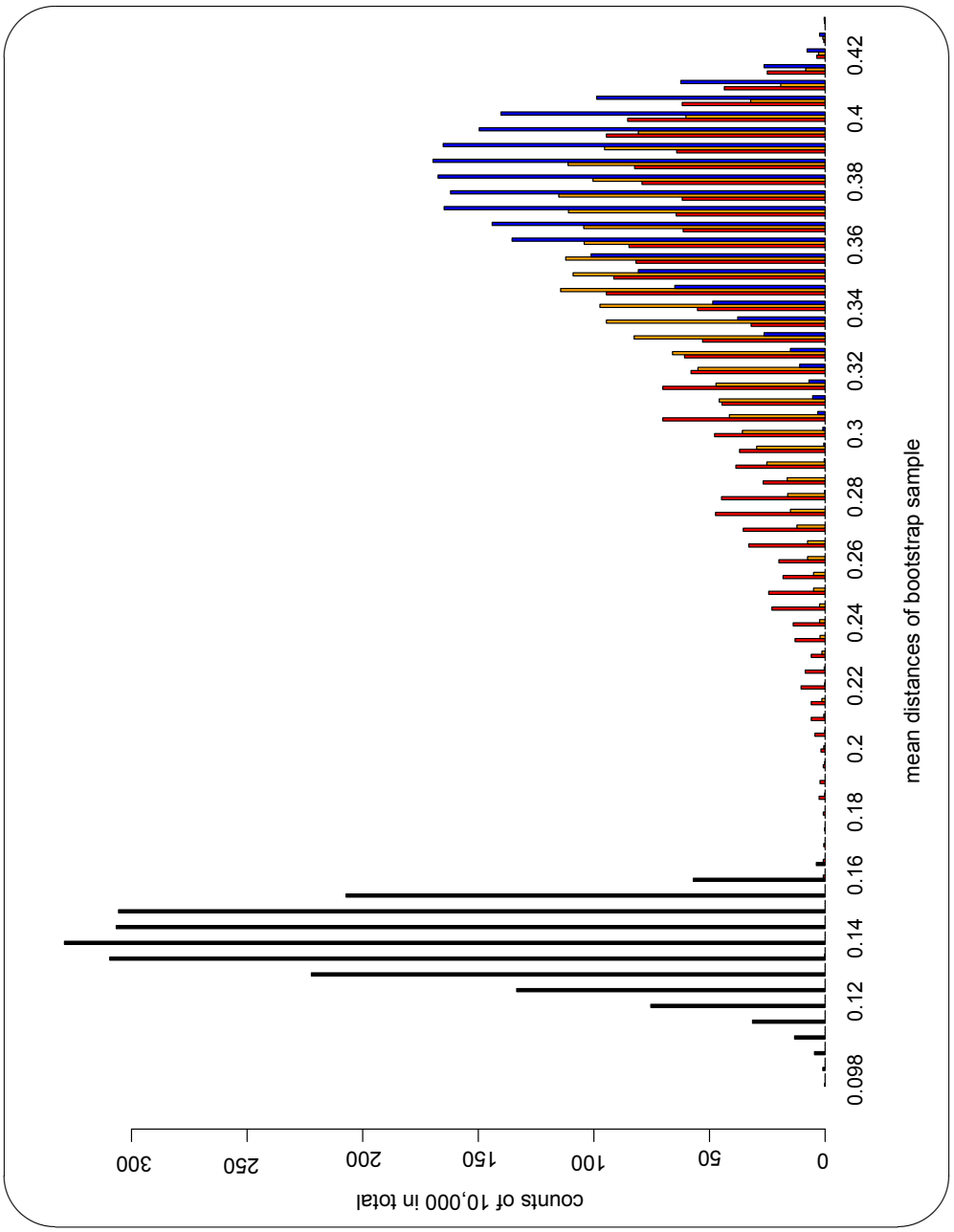


Fig. A.6. Distribution of mean distances between species from bootstrap samples for the different eubacterial phyla

Additional File 1.6 - Phylogenetic tree and distance matrix of Spirochaetes

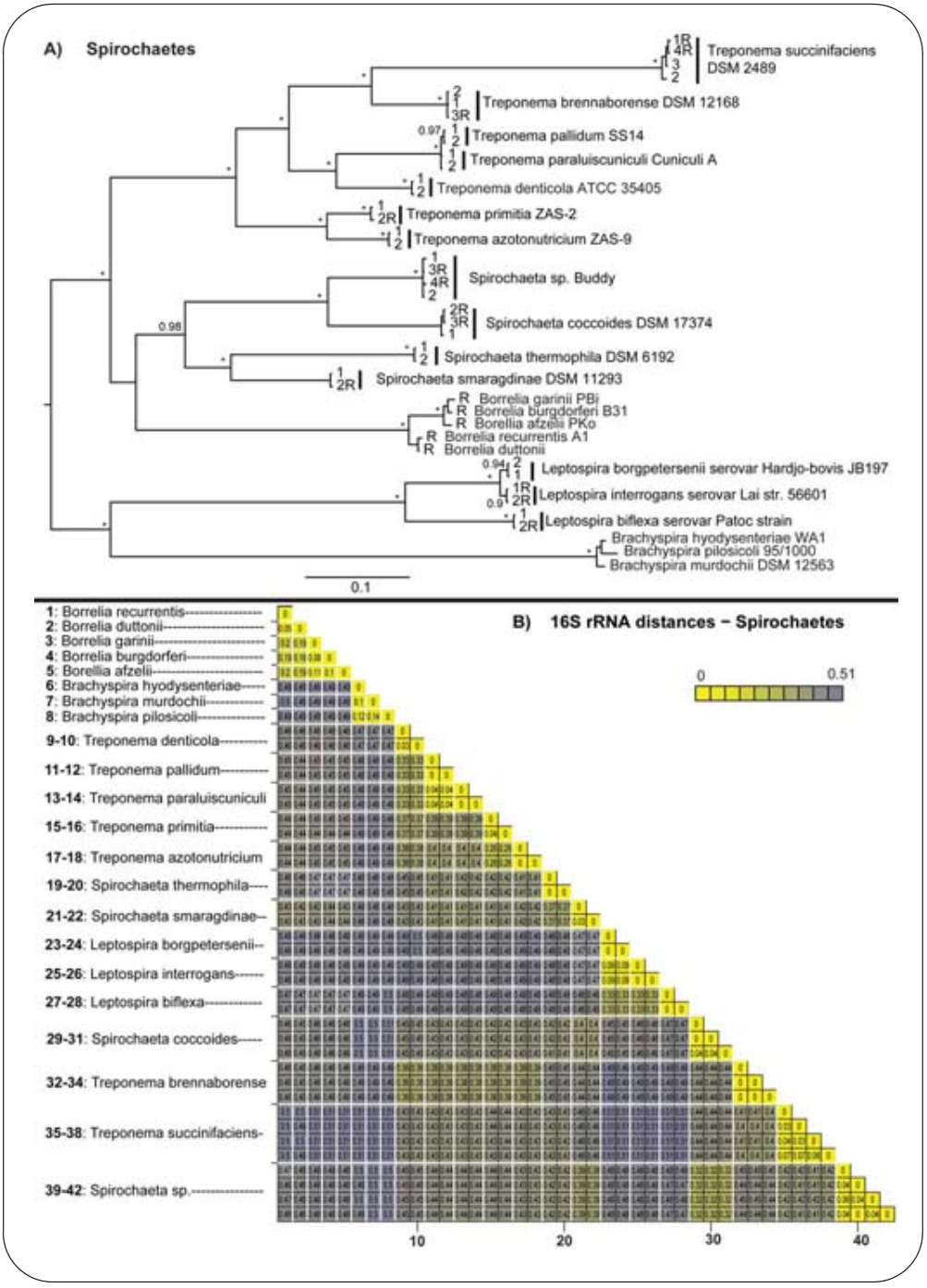
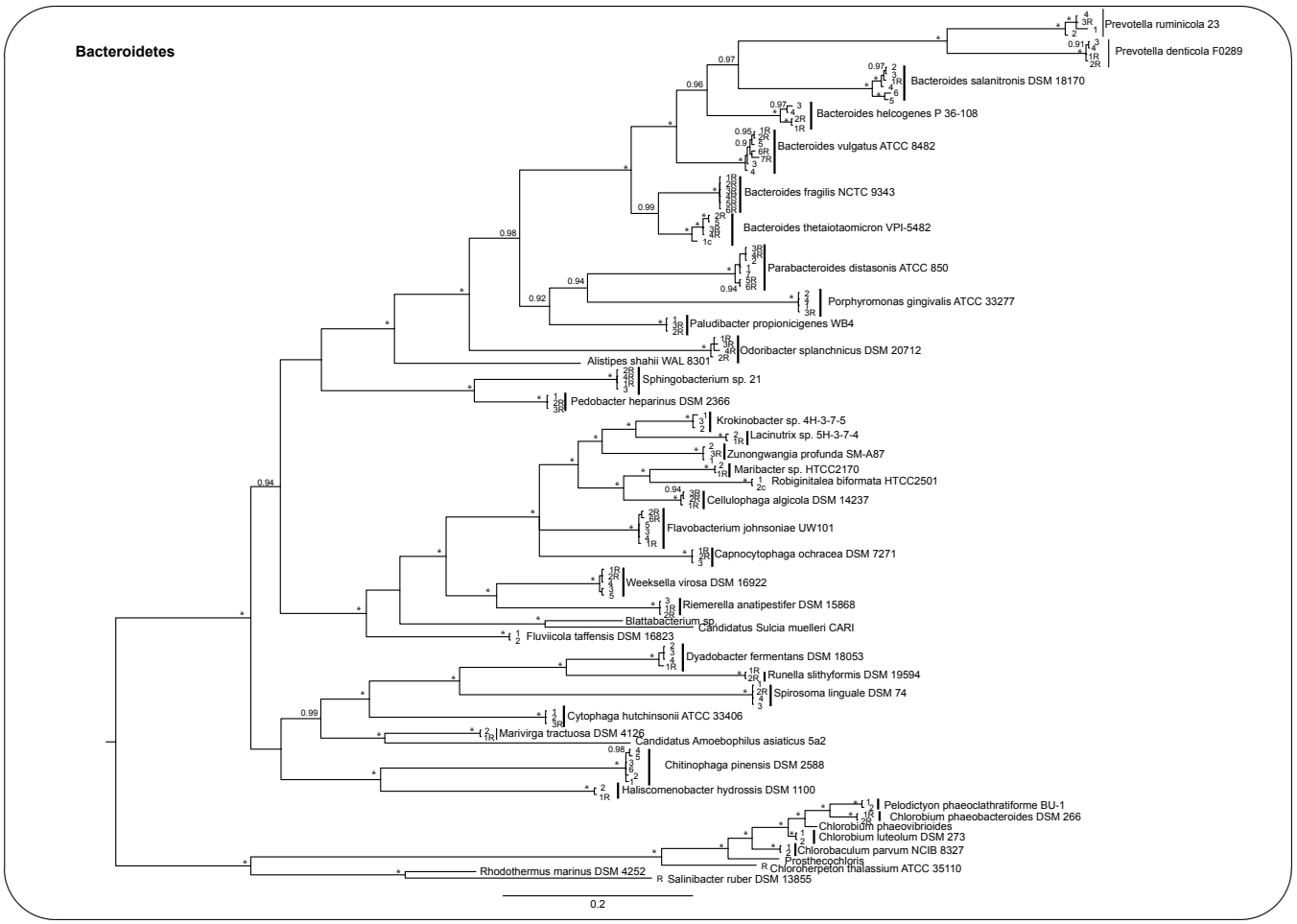


Fig. A.7. Phylogenetic tree and distance matrix of Spirochaetes(**A**) Phylogenetic tree of the eubacterial phylum Spirochaetes including all 16S rRNA gene copies, reconstructed using Bayesian analysis. On the nodes posterior probabilities > 0.90 are displayed. The letter "R" denote gene copies that are positioned on the reverse DNA strand. (**B**) Distance matrix of Spirochaetes. Genetic distances have been estimated according to the K80 substitution model. White lines separate sequence copies of different species.

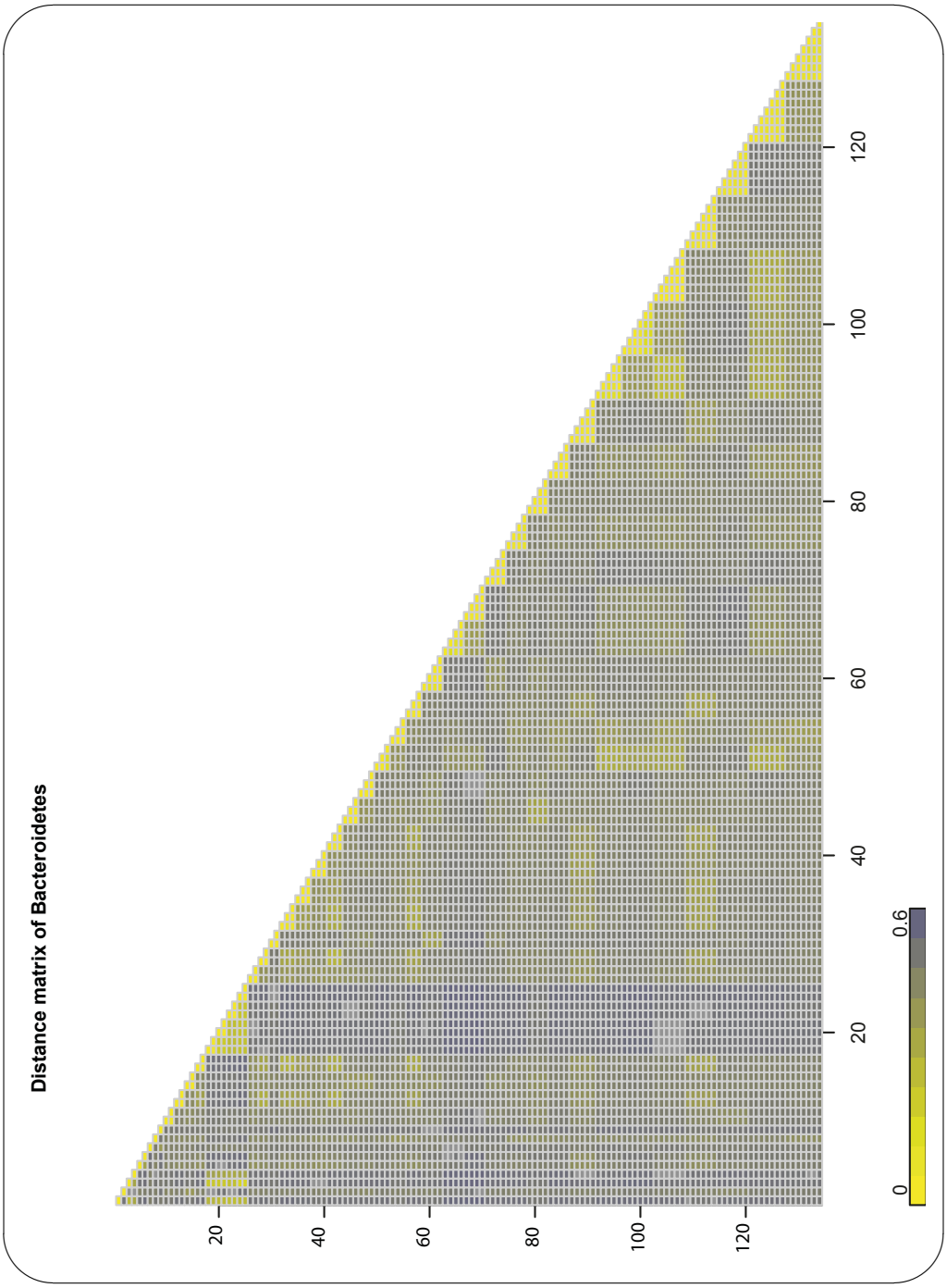
Additional File 1.7 - Phylogenetic tree of Bacteroidetes

Phylogenetic tree of the eubacterial phylum Bacteroidetes including all 16S rRNA gene copies, reconstructed using Bayesian analysis. On the nodes posterior probabilities > 0.90 are displayed.The letter “R" denote gene copies that are positioned on the reverse DNA strand.



Additional File 1.8 - Distance matrix of Bacteroidetes

Genetic distances have been estimated according to the K80 substitution model. White lines separate sequence copies of different species.



Additional File 1.9 - Distance matrix of cyanobacterial ITS-region

Distance matrix of the internal transcribed spacer sequence region in cyanobacteria. Genetic distances have been estimated according to the K80 substitution model. White lines separate sequence copies of different species. Distances ≥ 5.7 are displayed by the same blue color.

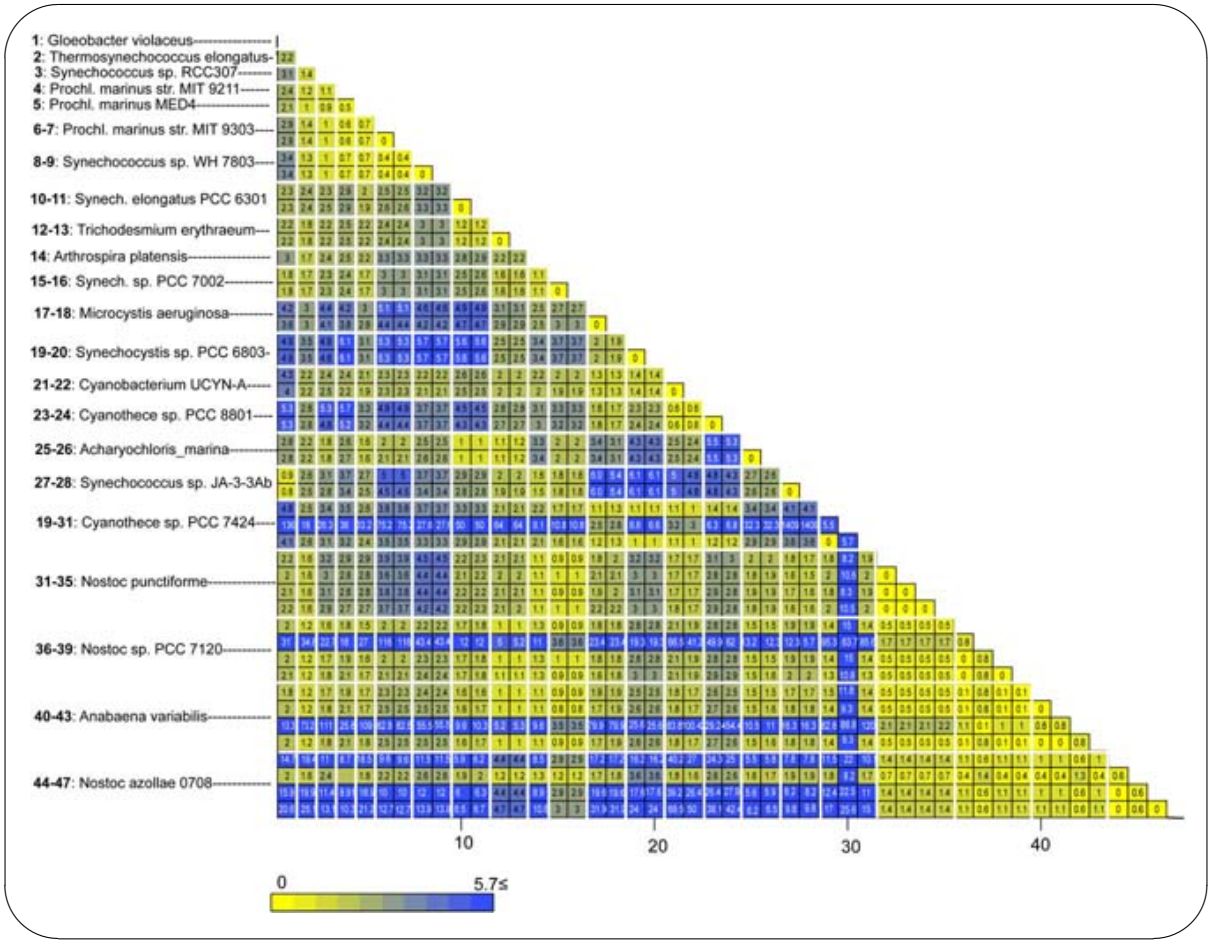


Table 0.1. Additional File 10 - Data of 16S rRNA gene sequences of the different eubacterial phyla. Species nomenclature, genome sizes, 16S rRNA gene copy numbers and accession numbers from the eubacterial taxa used in this study.

species	genome size	16S rRNA copy numbers	accession numbers
Aquificae			
Aquifex aeolicus VF5	1.6	2	AE000657.1
Desulfurobacterium thermolithotrophum DSM 11699	1.5	2	CP002543.1
Hydrogenobacter thermophilus TK-6	1.7	1	AP011112.1
Hydrogenobaculum sp. Y04AAS1	1.6	2	CP001130.1
Sulfurihydrogenibium sp. YO3AOP1	1.8	3	CP001080.1
Thermocrinis albus DSM 14484	1.5	1	CP001931.1
Thermovibrio ammonificans HB-1	1.8	3	CP002444.1
Chloroflexi			
Anaerolinea thermophila UNI-1	3.5	2	AP012029.1
Chloroflexus aggregans DSM 9485	4.7	3	CP001337.1
Chloroflexus aurantiacus J-10-fl	5.3	3	CP000909.1
Chloroflexus sp. Y-400-fl	5.3	3	CP001364.1
Dehalococcoides sp. BAV1	1.3	1	CP00688.1
Dehalococcoides ethenogenes 195	1.5	1	CP000027.1
Dehalogenimonas lykanthroporepellens BL-DC-9	1.7	1	CP002084.1
Herpetosiphon aurantiacus ATCC 23779	6.8	5	CP000875.1
Roseiflexus sp. RS-1	5.8	2	CP000686.1
Roseiflexus castenholzii DSM 13941	5.7	2	CP000804.1
Sphaerobacter thermophilus DSM 20745	4.0	1	CP001823.1
Thermomicrobium roseum DSM 5159	2.9	1	CP001275.1
Spirochaetae			
Borellia afzelii PKo	1.23783	1	CP000395.1
Borrelia burgdorferi B31	1.51913	1	AE000783.1
Borrelia duttonii Ly	1.567	1	CP000976.1
Borrelia garinii PBi	0.982668		CP000013.1
Borrelia hermsii DAH	1.059 5	1	CP000048.1
Borrelia recurrentis A1	0.930981	1	CP000993.1
Borrelia turicatae 91E135	0.92	1	CP000049.1
Brachyspira hyodysenteriae WA1	3.036	1	CP001357.1
Brachyspira murdochii DSM 12563	3.2	1	CP001959.1
Brachyspira pilosicoli 95/1000	2.6	1	CP002025.1
Leptospira biflexa serovar Patoc strain	3.603977	2	CP000777.1
Leptospira borgpetersenii serovar Hardjo-bovis JB197	3.87647	2	CP000350.1
Leptospira interrogans serovar Lai str. 56601	4.69224	2	AE010300.2
Spirochaeta coccoides DSM 17374	2.2	3	CP002659.1
Spirochaeta smaragdinae DSM 11293	4.7	2	CP002116.1
Spirochaeta sp. Buddy	3.3	4	CP002541.1
Spirochaeta thermophila DSM 6192	2.5	2	CP001698.1
Treponema azotonutricium ZAS-9	3.855671	2	CP001841.1
Treponema brennaborense DSM 12168	3.1	3	CP002696.1
Treponema denticola ATCC 35405	2.84	2	AE017226.1
Treponema pallidum subsp. pallidum SS14	1.14	2	CP000805.1
Treponema paraluiscuniculi Cuniculi A	1.1	2	CP002103.1
Treponema primitia ZAS-2	4.1	2	CP001843.1
Treponema succinifaciens DSM 2489	2.87	4	CP002631.1
Bacteroidetes			
Alistipes shahii WAL 8301	3.8	1	FP929032
Bacteroides fragilis NCTC 9343	5.23565	6	CR626927.1
Bacteroides helcogenes P 36-108	4	4	CP002352.1
Bacteroides salanitronis DSM 18170	4.2	6	CP002530.1
Bacteroides thetaiotaomicron VPI-5482	6.29336	5	AE015928.1
Bacteroides vulgatus ATCC 8482	5.16319	7	CP000139.1
Blattabacterium sp. (Blattella germanica) str. Bge	0.64	1	CP001487.1
Candidatus Amoebophilus asiaticus 5a2	1.9	1	CP001102.1
Candidatus Sulcia muelleri CARI	0.28	1	CP002163.1
Capnocytophaga ochracea DSM 7271	2.6	3	CP001632.1
Cellulophaga algicola DSM 14237	4.9	3	CP002453.1
Chitinophaga pinensis DSM 2588	9.12735	6	CP001699.1
Chlorobaculum parvum NCIB 8327	2.289249	2	CP001099.1
Chlorobium luteolum DSM 273	2.364842	2	CP000096.1
Chlorobium phaeobacteroides DSM 266	3.13390	2	CP000492.1
Chlorobium phaeovibrioides DSM 265	1.966858	1	CP000607.1
Chlorobium tepidum TLS	2.154946		AE006470.1
Chloroherpeton thalassium ATCC 35110	3.293456	1	CP001100.1
Cytophaga hutchinsonii ATCC 33406	4.433218	3	CP000383.1
Dyadobacter fermentans DSM 18053	6.96779	4	CP001619.1
Flavobacterium johnsoniae UW101	6.096872	6	CP000685.1
Fluviicola taffensis DSM 16823	4.633577	2	CP002542.1
Gramella forsetii KT0803	3.798465		CU207366.1
Haliscomenobacter hydrossis DSM 1100	8.371686	2	CP002691.1
Krokinobacter sp. 4H-3-7-5	3.389993	3	CP002528.1
Lacinutrix sp. 5H-3-7-4	3.296168	2	CP002825.1
Leadbetterella byssophila DSM 17132	4.059653		CP002305.1
Maribacter sp. HTCC2170	3.868304	2	CP002157.1
Marivirga tractuosa DSM 4126	4.511574	2	CP002349.1
Odoribacter splanchnicus DSM 20712	4.392288	4	CP002544.1
Paludibacter propionicigenes WB4	3.685504	3	CP002345.1
Parabacteroides distasonis ATCC 850	4.811379	7	CP000140.1
Pedobacter heparinus DSM 2366	5.167383	3	CP001681.1
Pelodictyon phaeoclathratiforme BU-1	3.018238	2	CP001110.1
Porphyromonas gingivalis ATCC 33277	2.354886	4	AP009380.1
Prevotella denticola F0289	2.937589	4	CP002589.1
Prevotella ruminicola 23	3.61955 9	4	CP002006.1
Prosthecochloris aestuarii DSM 271	2.512923	1	CP001108.1
Rhodothermus marinus DSM 4252	3.261604	1	CP001807.1
Riemerella anatipestifer DSM 15868	2.155121	3	CP002346.1
Robiginitalea biformata HTCC2501	3.530383	2	CP001712.1
Runella slithyformis DSM 19594	6.568739	2	CP002859.1
Salinibacter ruber DSM 13855	3.551823	1	CP000159.1
Sphingobacterium sp. 21	6.226409	4	CP002584.1
Spirosoma linguale DSM 74	8.078757	4	CP001769.1
Weeksella virosa DSM 16922	2.272954	5	CP002455.1
Zunongwangia profunda SM-A87	5.128187	3	CP001650.1

Additional Files to chapter II: The origin of multicellularity in cyanobacteria.

Additional File 2.1 - Taxon names of the phylogenetic tree of cyanobacteria (following pages)

Species names used in the phylogenetic analysis conducted with RAxML software. Taxon names are ordered by sub-groups as in Figure 1.

Clade D1 (Anabaena-Nostoc):

Anabaena circinalis AWQC310F 1295 bp, Aphanizomenon flos aquae A8 1295 bp, Anabaena smithii TAC450 1295 bp, Aphanizomenon flos aquae strAph Zayi 1295 bp, Anabaena circinalis 1tu33s12 r strain 1tu33s12 1295 bp, Anabaena ellipsoides Ana HB 1295 bp, Anabaena planctonica TAC424 1295 bp, Anabaena crassa CENA207 1295 bp, Anabaena circinalis AWT205B 1295 bp, Anabaena crassa CENA196 1295 bp, Anabaena circinalis AWQC331C 1295 bp, Anabaena circinalis AWQC332H 1295 bp, Anabaena crassa CENA206 1295 bp, Anabaena flos aquae 04 53 r strain 04 53 1295 bp, Anabaena affinis NIES 40 1295 bp, Anabaena spiroides r strain PMC9702 1295 bp, Anabaena spiroides NIES 79 1295 bp, Anabaena sigmoidea 0tu36s7 r strain 0tu36s7 1295 bp, Anabaena sp1 strAna Ku5 1295 bp, Anabaena planctonica CENA209 1295 bp, Aphanizomenon flos aquae 1tu37s13 r strain 1tu37s13 1295 bp, Aphanizomenon flos aquae AFA 3 1295 bp, Aphanizomenon flos aquae r strain PMC9707 1295 bp, Anabaena sigmoidea 0tu38s4 r strain 0tu38s4 1295 bp, Aphanizomenon flos aquae A7 1295 bp, Aphanizomenon flos aquae 1tu29s19 r strain 1tu29s19 1295 bp, Aphanizomenon flos aquae strAph Ku 1295 bp, Anabaena viguieri TAC433 1295 bp, Aphanizomenon flos aquae 617 1295 bp, Anabaena smithii TAC431 1295 bp, Aphanizomenon flos aquae A1 1295 bp, Aphanizomenon flos aquae AFA 6 1295 bp, Aphanizomenon flos aquae A5 1295 bp, Anabaena solitaria BC Ana 0025 1295 bp, Anabaena flos aquae EH 1 1295 bp, Anabaena flos aquae DC 1 1295 bp, Anabaena spiroides 1tu39s17 r strain 1tu39s17 1295 bp, Anabaena smithii 1tu39s8 r strain 1tu39s8 1295 bp, Anabaena planctonica NIVA CYA 66 1295 bp, Aphanizomenon flos aquae strAph Inba 1295 bp, Anabaena ucrainica TAC449 1295 bp, Anabaena circinalis NIES41 1295 bp, Anabaena spiroides NIES 76 1295 bp, Anabaena circinalis CENA193 1295 bp, Anabaena planctonica TAC421 for 1295 bp, Aphanizomenon flos aquae strAph K2 1295 bp, Anabaena circinalis CENA191 1295 bp, Aphanizomenon flos aquae varKlebahnii r strain 218 1295 bp, Anabaena flos aquae DC 2 1295 bp, Anabaena flos aquae AWQC112D 1295 bp, Anabaena planctonica TAC435 1295 bp, Anabaena circinalis AWQC307C 1295 bp, Aphanizomenon flos aquae r strain PMC9401 1295 bp, Aphanizomenon flos aquae 1tu26s2 r strain 1tu26s2 1295 bp, Anabaena circinalis CENA190 1295 bp, Anabaena planctonica Inba2 1295 bp, Anabaena circinalis AWT001 1295 bp, Aphanizomenon spTR183 r strain TR183 1295 bp, Anabaena spBIR41 1295 bp, Anabaena flos aquae 0tu33s2a r strain 0tu33s2a 1295 bp, Anabaena spBIR259 1295 bp, Aphanizomenon gracile UADFA16 1295 bp, Anabaena spXPORK15F 1295 bp, Aphanizomenon flos aquae strain NIES81 1295 bp, Anabaena spA277 r strain A277 1295 bp, Anabaena sp18B6 1295 bp, Anabaena mendotae 04 11 r strain 04 11 1295 bp, Aphanizomenon gracile strAph NH 5 1295 bp, Anabaena spBIR374 1295 bp, Anabaena spBIR246 1295 bp, Aphanizomenon issatschenkoi strTAC419 1295 bp, Anabaena flos aquae NIVA CYA83/1 r strain CYA83/1 1295 bp, Aphanizomenon gracile strain LMECYA40 1295 bp, Anabaena spBIR202 1295 bp, Anabaena sp1tu34s7 r strain 1tu34s7 1295 bp, Aphanizomenon cfgracile 271 r strain 271 1295 bp, Anabaena lemmermannii strAna Dalai 1295 bp, Anabaena sp318 1295 bp, Anabaena spBIR52 1295 bp, Anabaena spBIR358 1295 bp, Aphanizomenon gracile UADFA11 1295 bp, Anabaena spBIR76 1295 bp, Aphanizomenon flos aquae r strain PMC9706 1295 bp, Anabaena lemmermannii 04 42 r strain 04 42 1295 bp, Aphanizomenon flos aquae 1295 bp, Anabaena oscillarioides BECID22 r strain BECID22 1295 bp, Aphanizomenon issatschenkoi 1313 1295 bp, Anabaena cfcylindrica PMC9705 r strain PMC9705 1295 bp, Anabaena flos aquae 04 40 r strain 04 40 1295 bp, Aphanizomenon issatschenkoi 2312 1295 bp, Anabaena solitaria r strain 82 1295 bp, Anabaena flos aquae 1tu35s12 r strain 1tu35s12 1295 bp, Aphanizomenon spBC Aph 9601 r 1295 bp, Anabaena spBIR348 1295 bp, Anabaena lemmermannii 1tu32s11 r strain 1tu32s11 1295 bp, Anabaena sp315 1295 bp, Anabaena spBIR30 1295 bp, Anabaena flos aquae strain PCC 9302 1295 bp, Anabaena flos aquae 0tu33s15 r strain 0tu33s15 1295 bp, Aphanizomenon issatschenkoi strLEMCYA31 1295 bp, Anabaena sp66A r strain 66A 1295 bp, Anabaena Lemmermannii r strain 202A2 1295 bp, Anabaena lemmermannii BC Ana 0005 1295 bp, Anabaena cffallax CENA208 1295 bp, Anabaena sp90 r strain 90 1295 bp, Aphanizomenon issatschenkoi 473 1295 bp, Anabaena spBIR370B 1295 bp, Anabaena spBIR256 1295 bp, Anabaena lemmermannii TAC437 1295 bp, Anabaena spBIR49 1295 bp, Aphanizomenon gracile 1tu26s16 r strain 1tu26s16 1295 bp, Aphanizomenon gracile 1040 1295 bp, Aphanizomenon flos aquae r strain NIES81 1295 bp, Anabaena sp0tu39s7 r strain

0tu39s7 1295 bp, Anabaena planctonica r strain 71 1295 bp, Anabaena flos aquae 1tu30s4 r strain 1tu30s4 1295 bp, Anabaena spBIR162 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51018 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51006 1295 bp, Cylindrospermopsis raciborskii strain Germany 1 1295 bp, Cylindrospermopsis raciborskii BM 1295 bp, Anabaena spXP6B 1295 bp, AF067819 Cylindrospermopsis raciborskii form 2 1295 bp, Cylindrospermopsis raciborskii strain 05E 1295 bp, Cylindrospermopsis raciborskii for r strain CRJ1 1295 bp, Cylindrospermopsis raciborskii strain Brazil 1 1295 bp, Anabaena spBIR361 1295 bp, Anabaena cylindrica XP6B r strain XP6B 1295 bp, Cylindrospermopsis raciborskii strain Germany 2 1295 bp, Cylindrospermopsis raciborskii strain Mk 1295 bp, Anabaena reniformis 07 01 r isolate 07 01 1295 bp, Cylindrospermopsis raciborskii strain LJ 1295 bp, Cylindrospermopsis raciborskii QHSS/NR/CYL/03 1295 bp, Anabaena spBIR5 1295 bp, Cylindrospermopsis raciborskii strain Marau 1 1295 bp, Anabaena spBIR440 1295 bp, Anabaena cylindrica 1295 bp, Anabaena spBECID20 1295 bp, Anabaena spXSP2A 1295 bp, Anabaena spXP34A 1295 bp, Cylindrospermopsis raciborskii strain Florida I 1295 bp, Cylindrospermopsis raciborskii strain Bal 6 1295 bp, Cylindrospermopsis raciborskii strain 23D 1295 bp, Wollea saccata ACCS 045 1295 bp, Cylindrospermopsis raciborskii strain Florida G 1295 bp, Trichormus variabilis HINDAK 2001/4 r strain HINDAK 2001/4 1295 bp, Cylindrospermopsis raciborskii strain Aqc 1295 bp, Raphidiopsis mediterranea HB2 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51016 1295 bp, Nostoc spAzolla cyanobiont small subunit 1295 bp, Anabaena flos aquae NRC525 17 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51004 1295 bp, Anabaena spBIR84 1295 bp, Anabaena spBIR272 1295 bp, Anabaena compacta CCAP 1403/24 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51019 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51015 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51009 1295 bp, Cylindrospermopsis raciborskii strain 09A 1295 bp, Cylindrospermopsis raciborskii strain Caia 1295 bp, Anabaena flos aquae NIES 73 1295 bp, AF317629 Anabaena spPCC 7108 1295 bp, Anabaena spXSP36B 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51017 1295 bp, Anabaena oscillarioides BO HINDAK 1984/43 r strain BO HINDAK 1984/43 1295 bp, Cyanobacterium BECID34 r strain BECID34 1295 bp, Anabaena compacta ANACOM KOR r strain ANACOM KOR 1295 bp, Anabaena cfcylindrica 133 r strain 133 1295 bp, Anabaena spWH School stisolate for r 1295 bp, Anabaena spBIR169 1295 bp, Cylindrospermopsis raciborskii for r strain NIES991 1295 bp, Anabaena spXP15D 1295 bp, Anabaena spBECID23 1295 bp, Anabaena spBIR25 1295 bp, Cylindrospermopsis raciborskii FAS C1 1295 bp, Cylindrospermopsis raciborskii 1295 bp, Cylindrospermopsis raciborskii Goon 1295 bp, Anabaena spXP36D 1295 bp, Anabaena spXP6A 1295 bp, Anabaena spXP6C 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51003 1295 bp, Anabaena spSSM 00 1295 bp, Cylindrospermopsis raciborskii for r strain NIES993 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51010 1295 bp, Cylindrospermopsis raciborskii strain Florida F 1295 bp, Anabaena cylindrica DC 3 1295 bp, Anabaena spBIR96 1295 bp, Cylindrospermopsis raciborskii strain 23B 1295 bp, Anabaena sp PCC 7108 1295 bp, Cylindrospermopsis raciborskii strain Florida D 1295 bp, Cylindrospermopsis raciborskii strain Brazil 2 1295 bp, Cylindrospermopsis raciborskii for r strain DMKU51013 1295 bp, Anabaena spXPORK27C 1295 bp, Cylindrospermopsis raciborskii strain 4799 1295 bp, Anabaena spBIR2 1295 bp, Cylindrospermopsis raciborskii strain 24C 1295 bp, Anabaena spKVJF17 1295 bp, Cylindrospermopsis raciborskii for r strain NIES992 1295 bp, Aphanizomenon aphanizomenoides 04 43 r isolate 04 43 1295 bp, Raphidiopsis curvata HB1 1295 bp, Anabaena sedovii ACCS 058 1295 bp, Cylindrospermopsis raciborskii T3 1295 bp, Anabaena aphanizomenoides CENA188 1295 bp, Anabaena flos aquae strain AWQC 264A 1295 bp, Nodularia sphaerocarpa r strain UTEX B 2093 1295 bp, Anabaenopsis spPCC 9215 1295 bp, Nodularia spumigena strain L575 1295 bp, Nodularia spumigena r strain AV1 1295 bp, Nodularia harveyana r strain Bo53 1295 bp, Nodularia sphaerocarpa r strain BECID35 1295 bp, Nodularia spumigena r strain NSPI 05 1295 bp, Nodularia spumigena strain NSOR12 1295 bp, Nodularia harveyana r strain BECID29 1295 bp, Nodularia spumigena strain PCC73104 1295 bp, Nodularia spumigena strain NSKR07 1295 bp, Nodularia sphaerocarpa r strain Up16a 1295 bp, Nodularia sphaerocarpa r strain PCC73104 1295 bp, Nodularia spumigena strain BY1 1295 bp, Calothrix spBECID18 r strain BECID18 1295 bp, Nodularia spumigena r strain HEM 1295 bp, Nodularia spumigena strain HEM 1295 bp, Anabaena cylindrica NIES19

1295 bp, Anabaenopsis abijatae AB2002/18 r strain AB2002/18 1295 bp, Anabaena spBECID19 1295 bp, Anabaena spXP35A 1295 bp, Nodularia spumigena strain UTEX B2092 1295 bp, Nodularia sphaerocarpa 1295 bp, Nodularia spWH Baltic Sea isolate for r 1295 bp, Anabaenopsis cfAB2002/25 abijatae r strain AB2002/25 1295 bp, Nodularia harveyana r strain Lukesova 18/94 1295 bp, Nodularia spumigena strain NSLA01 1295 bp, Nodularia sphaerocarpa PCC 7804 1295 bp, Nodularia spumigena for r 1295 bp, Anabaenopsis elenkinii AB2002/37 r strain AB2002/37 1295 bp, Nodularia harveyana strain PCC7804 1295 bp, Nodularia harveyana strain CDAC1983/300 1295 bp, Nodularia sphaerocarpa r strain BECID36 1295 bp, Anabaenopsis sp1A 1295 bp, Nodularia harveyana strain UTEX B2093 1295 bp, Nodularia spF81 1295 bp, Aphanizomenon spNH 5 1295 bp, Nodularia spumigena strain NSPH02 1295 bp, Nodularia spumigena strain NSBL05 1295 bp, Nodularia spumigena strain NSBR01 1295 bp, Nodularia spPCC 9350 1295 bp, Nodularia spumigena strain NSGL02A10 1295 bp, Cyanospira rippkae 1295 bp, Nodularia spumigena r strain NSOR 12 1295 bp, Nodularia spumigena r strain GR8b 1295 bp, Nodularia spumigena strain NSPH05A14 1295 bp, Anabaenopsis elenkinii AB2006/20 r strain AB2006/20 1295 bp, Anabaena spSKJF11 1295 bp, Anabaenopsis elenkinii AB2002/17 r strain AB2002/17 1295 bp, Nodularia sphaerocarpa r strain Fae19 1295 bp, Nodularia spumigena r strain Huebel 1988/306 1295 bp, Nodularia spumigena r strain F81 1295 bp, Nodularia spumigena r strain AV63 1295 bp, Nodularia spumigena strain NSLA02A4 1295 bp, Anabaena spBECID31 1295 bp, Nodularia spumigena strain HKVV 1295 bp, Nodularia spKAC 17 for r 1295 bp, Nodularia spLEGE06071 1295 bp, Anabaena augstumalis SCMIDKE JAHNKE/4a r strain SCMIDKE JAHNKE/4a 1295 bp, Nodularia spLukesova 1/91 r strain Lukesova 1/91 1295 bp, Anabaena spBECID8 1295 bp, Nodularia harveyana r strain BECID27 1295 bp, Nodularia spumigena GSL023 1295 bp, Anabaena spXPORK36C 1295 bp, Aphanizomenon ovalisporum FAS AP1 1295 bp, AF160256 Anabaena bergii 1295 bp, Umezakia natans TAC101 1295 bp, Nostoc spPeltigera membranacea 4 cyanobiont 1295 bp, Nostoc commune UTEX 584 1295 bp, AF062637 Nostoc spGSV224 1295 bp, Nostoc spLeptogium gelatinosum cyanobiont 1295 bp, Nostoc spUK18 1295 bp, AF027655 Nostoc PCC73102 1295 bp, Nostoc spMollenhauer 1 1 125 1295 bp, Nostoc sp1tu14s8 r strain 1tu14s8 1295 bp, Nostoc commune for r country Japan Ishikawa 1295 bp, Nostoc spDM103 r strain SAG 2028 1295 bp, Nostoc spPeltigera pruinosa cyanobiont 14 1295 bp, Nostoc spLobaria cyanobiont 34 1295 bp, Nostoc spPannaria cfallorhiza cyanobiont NZ 1295 bp, Nostoc spPannaria affleproloma cyanobiont 1a NZ 1295 bp, Nostoc sphaeroides HBHF0604 1295 bp, Nostoc spKVJF4 1295 bp, Nostoc spPeltigera canina 1 cyanobiont 1295 bp, Nostoc spNephroma laevigatum cyanobiont 39 1295 bp, Nostoc spPannaria sphinctrina cyanobiont 1b Ch 1295 bp, Nostoc spPannaria euphylla cyanobiont NZ 1295 bp, Nostoc spNephroma parile cyanobiont 32 1295 bp, Nostoc spLobaria hallii cyanobiont 19 1295 bp, Nostoc spPannaria affleproloma cyanobiont 2 Ch 1295 bp, Nostoc spSAG 29 90 1295 bp, Nostoc spPannaria conoplea cyanobiont No 1295 bp, AF062638 Nostoc ATCC53789 1295 bp, Nostoc commune AHNG0605 1295 bp, Nostoc spMollenhauer 1 1 108 1295 bp, Gloeotrichia echinulata URA3 r environmental colony URA3 1295 bp, Nostoc spSKS2 1295 bp, Nostoc commune for r strain KU002 1295 bp, Nostoc sp9104 r strain 9104 1295 bp, Nostoc punctiforme SAG 71 79 1295 bp, Nostoc spPannaria rubiginella cyanobiont Chile 1295 bp, Nostoc spPeltigera canina 2 cyanobiont 1295 bp, Nostoc spSAG 39 87 1295 bp, Nostoc spMuscicolous cyanobiont 21 1295 bp, Nostoc spPeltigera lepidophora cyanobiont 1295 bp, Nostoc flagelliforme IMGA0408 1295 bp, Nostoc muscorum SAG 57 79 1295 bp, Nostoc spMollenhauer 1 1 115 1295 bp, Nostoc sp8964 3 r strain 8964 3 1295 bp, Nostoc spPannaria mosenii cyanobiont Mex 1295 bp, Nostoc spSKS5 1295 bp, Nostoc punctiforme SAG 69 79 isolation source lichen specimen voucher SAG 69 79 1295 bp, Nostoc spNephroma parile cyanobiont 27 1295 bp, Nostoc spSKJ4 1295 bp, Nostoc entophytum IAM M 267 for 1295 bp, Nostoc spPeltigera didactyla 3 cyanobiont 1295 bp, Nostoc spLobaria pulmonaria cyanobiont 36 1295 bp, Nostoc spPannaria andina 2 cyanobiont Chile 1295 bp, Nostoc spMollenhauer 94 1 1295 bp, Nostoc spPannaria affsphinctrina cyanobiont 1a NC 1295 bp, Nostoc spSKJ1 1295 bp, Nostoc spPannaria araneosa cyanobiont NZ 1295 bp, Nostoc spPannaria fulvescens cyanobiont NZ 1295 bp, Nostoc spKVJF15 1295 bp, Nostoc spPannaria patagonica cyanobiont Ch 1295 bp, Nostoc spNephroma helveticum cyanobiont 1295 bp, Nostoc commune for r country Japan Yamaguchi 1295 bp, Nostoc sp8963 small subunit 1295 bp, Nostoc spSKSF2 1295 bp, Nostoc spPannaria athroophylla

cyanobiont Ch 1295 bp, Nostoc spKVSF4 1295 bp, Nostoc commune for r country Japan Okinawa Ryukyu University 1295 bp, Nostoc sp0GU36S01 r strain 0GU36S01 1295 bp, Nostoc calcicola VI r strain VI 1295 bp, Nostoc spPeltigera rufescens 3 cyanobiont 1295 bp, AF317630 Nostoc spAWT 203 1295 bp, Nostoc spKVS1 1295 bp, Nostoc spSKJF2 1295 bp, Nostoc spMollenhauer 1 1 066 1295 bp, Nostoc spPeltigera membranacea 5 cyanobiont 1295 bp, Nostoc spMollenhauer 1 1 106b2 1295 bp, Nostoc spPeltigera rufescens 2 cyanobiont 1295 bp, Nostoc spSticta beauvoisii cyanobiont 1295 bp, Nostoc cfcommune KG 54 for r 1295 bp, Nostoc spNephroma helveticum cyanobiont 37 1295 bp, Nostoc commune for r country Japan Hokkaido 1295 bp, Nostoc commune for r country Japan Saga 1295 bp, Nostoc sp195 A21 1295 bp, Nostoc spKVJF1 1295 bp, Nostoc spPannaria affleproloma cyanobiont 1b NZ 1295 bp, Nostoc spPeltigera canina 4 cyanobiont 1295 bp, Nostoc commune for r 1295 bp, Nostoc spPannaria andina 1 cyanobiont Chile 1295 bp, Nostoc spSKSF3 1295 bp, Nostoc spPeltigera degeni cyanobiont 1295 bp, Nostoc spPCC 9229 small subunit 1295 bp, Nostoc spKVJ20 1295 bp, Nostoc spKVJ2 1295 bp, Nostoc spSKS9 1295 bp, Nostoc punctiforme SAG 68 79 isolation source lichen specimen voucher SAG 68 79 1295 bp, Nostoc spPannaria obscura cyanobiont Aus 1295 bp, Nostoc spPannaria isabellina cyanobiont 2 Ch 1295 bp, Nostoc spSKSL2 1295 bp, Nostoc spPeltigera collina cyanobiont 20 1295 bp, Nostoc spPannaria andina cyanobiont Peru 1295 bp, Gloeotrichia echinulata PYH14 r environmental colony PYH14 1295 bp, Nostoc sp17 1295 bp, Nostoc spNi4 C1 for r 1295 bp, Nostoc spPeltigera membranacea 1 cyanobiont 1295 bp, Nostoc spKVJF8 1295 bp, Nostoc spLukesova 5/96 r strain Lukesova 5/96 1295 bp, Nostoc spLukesova 40/93 r strain Lukesova 40/93 1295 bp, Nostoc spPannaria elixii cyanobiont 2 NZ 1295 bp, Nostoc spMollenhauer 1 1 150b 1295 bp, Nostoc spPeltigera pruinosa cyanobiont 18 1295 bp, Nostoc sp195 A22 1295 bp, Nostoc spCollema crispum cyanobiont 1295 bp, Nostoc spLobaria hallii cyanobiont 1295 bp, Nostoc spNephroma bellum cyanobiont 1295 bp, Nostoc spMassalongia carnosa cyanobiont 1295 bp, Nostoc flagelliforme strSunitezuoqi 1295 bp, Nostoc spAl3 r strain Al3 1295 bp, Nostoc spParmeliella triptophylla cyanobiont 30 1295 bp, Nostoc spPCC 9305 small subunit 1295 bp, Nostoc spSAG 41 87 1295 bp, Nostoc spPannaria durietzii cyanobiont 1 NZ 1295 bp, Nostoc commune 0Brien 02011101 1295 bp, Nostoc spPeltigera rufescens 4 cyanobiont 1295 bp, Nostoc sp8901 1 r strain 8901 1 1295 bp, Nostoc commune for r strain SO 42 1295 bp, Nostoc spSKJ2 1295 bp, Nostoc spPeltigera venosa cyanobiont 15 1295 bp, Nostoc sp8926 r strain 8926 1295 bp, Nostoc spNephroma helveticum cyanobiont 33 1295 bp, Nostoc commune for country Japan Hyogo Akashi 1295 bp, Nostoc spIO 102 I 1295 bp, Nostoc edaphicum X r strain X 1295 bp, Nostoc spSAG 36 92 1295 bp, Nostoc spPeltigera neopolydactyla cyanobiont 1295 bp, Nostoc spKVJF16 1295 bp, Nostoc spSticta fuliginosa cyanobiont 1295 bp, Nostoc commune for r country France Grenoble 1295 bp, Nostoc spHKAR 2 1295 bp, Nostoc spOs1 C1 for r 1295 bp, Anabaena flos aquae RPAN52 clone 1 1295 bp, Anabaena spCH1 1295 bp, Nostoc spCc2 r strain Cc2 1295 bp, Nostoc spCENA105 1295 bp, Trichormus azollae Kom BAI/1983 r strain Kom BAI/1983 1295 bp, Nostoc spPCC 9231 small subunit 1295 bp, Nostoc piscinale BF3 1295 bp, Nostoc spMau15 r strain Mau15 1295 bp, Nostoc spPCC 7906 for r 1295 bp, Anabaena iyengarii RPAN70 clone 1 1295 bp, Anabaena azotica r 1295 bp, Nostoc spPCC 7423 1295 bp, Nostoc spTO1S01 r strain TO1S01 1295 bp, Nostoc spKK 01 for 1295 bp, Anabaena spiroides RPAN57 clone 1 1295 bp, Anabaena iyengarii RPAN6 clone 1 1295 bp, Nostoc spDe1 r strain De1 1295 bp, Nostoc spPCC 9426 r strain PCC 9426 1295 bp, Nostoc sp8941 small subunit 1295 bp, Nostoc sp8916 small subunit 1295 bp, Nostoc linckia vararvense IAM M 30 for r 1295 bp, Anabaena flos aquae UTCC 64 1295 bp, Nostoc ellipsosporum V r strain V 1295 bp, Nostoc sp8938 small subunit 1295 bp, Calothrix spMk1 C1 for r 1295 bp, Anabaena variabilis RPAN16 clone 1 1295 bp, Anabaena sp r 1295 bp, Nostoc spCr4 r strain Cr4 1295 bp, Nostoc muscorum CENA61 1295 bp, Anabaena variabilis for r 1295 bp, Nostoc spPCC 7120 1295 bp, Nostoc spTH1S01 r strain TH1S01 1295 bp, Cylindrospermum spA1345 1295 bp, Anabaena torulosa BF1 1295 bp, Nostoc spYK 01 for 1295 bp, Nostoc spPCC 8976 r strain PCC 8976 1295 bp, Nostoc spPCC 6720 1295 bp, Tolypothrix spIAM M 259 for 1295 bp, Cylindrospermum spPCC 7417 r strain PCC 7417 1295 bp, Nostoc calcicola 99 1295 bp, Nostoc spCENA107 1295 bp, Calothrix spPCC 7101 for r 1295 bp, Nostoc spPCC 8112 r strain PCC 8112 1295 bp, Calothrix elenkinii RPC1 1295 bp, Nostoc sp152 r strain 152 1295 bp, Nostoc spMollenhauer 1 1 067 1295 bp, Nostoc spHK 01 for 1295 bp, Pseudanabaena spMBIC10772 for r strain MBIC10772 1295 bp, Tolypothrix spPCC 7504 r strain PCC 7504

1295 bp, Nostoc spMollenhauer 1 1 088 1295 bp, Nostocaceae cyanobacterium SAG B11 82 1295 bp, Nostoc elgonense TH3S05 r strain TH3S05 1295 bp, Nostoc muscorum CENA18 1295 bp, Tolypothrix spTOL328 r strain TOL328 1295 bp, Tolypothrix spPCC 7415 r strain PCC 7415 1295 bp, Nostoc spKU001 for r 1295 bp, Nostoc spAl1 r strain Al1 1295 bp, Nostoc spCam2S01 r strain Cam2S01 1295 bp, Nostoc calcicola r strain TH2S22 1295 bp, Trichormus doliolum strdoliolum 1 r strain doliolum 1 1295 bp, Nostoc spSKS8 1295 bp, Nostoc verrucosum KU005 for r 1295 bp, Cylindrospermum spCENA33 1295 bp, Nostoc spCENA88 1295 bp, Nostoc carneum BF2 1295 bp, Nostoc carneum IAM M 35 for r 1295 bp, Cylindrospermum licheniforme UTEX 2014 for r 1295 bp

Clade D2 (Calothrix-Fischerella):

Calothrix spBECID30 r strain BECID30 1295 bp, Symphyonema sp1517 r strain 1517 1295 bp, Scytonema spU 3 3 1295 bp, Calothrix D253 r 1295 bp, Rivularia spXSP25A r strain XSP25A 1295 bp, Calothrix spBECID9 r strain BECID9 1295 bp, Rivularia spBECID10 r strain BECID10 1295 bp, Calothrix spBECID1 r strain BECID1 1295 bp, Calothrix spPCC 8909 r strain PCC 8909 1295 bp, Calothrix spBECID26 r strain BECID26 1295 bp, Nostochopsis lobatus 92 1 r strain 92 1 1295 bp, Rivularia spIAM M 261 for r 1295 bp, Calothrix spBIR LS5 r strain BIR LS5 1295 bp, Rivularia atra BIR MGR1 r environmental colony BIR MGR1 1295 bp, Fischerella spCENA161 1295 bp, Rivularia spXP16B r strain XP16B 1295 bp, Stigonema ocellatum SAG 48 90 r strain SAG 48 90 1295 bp, Brasilonema bromeliae SPC 951 1295 bp, Rivularia atra BIR KRIV1 r environmental colony BIR KRIV1 1295 bp, Fischerella thermalis PCC 7521 for r 1295 bp, Hapalosiphon spIAM M 264 for 1295 bp, Westiellopsis prolifica SAG 16 93 r strain SAG 16 93 1295 bp, Rivularia spXP3A r strain XP3A 1295 bp, Chlorogloeopsis spGreenland 5 1295 bp, Symphyonema sp1269 1 r strain 1269 1 1295 bp, Mastigocladopsis repens MORA r strain MORA 1295 bp, Chlorogloeopsis fritschii PCC 6912 for 1295 bp, Chroococcidiopsis spCC4 1295 bp, CfCalothrix spMuscicolous cyanobiont 5 1295 bp, Calothrix spBECID14 r strain BECID14 1295 bp, Brasilonema roberti lammi strlos manantiales1 1295 bp, Calothrix spPCC 7507 r strain PCC 7507 1295 bp, Calothrix spUKK3412 r strain UKK3412 1295 bp, Brasilonema octagenarum UFV OR1 1295 bp, Hapalosiphon hibernicus BZ 3 1 1295 bp, Fischerella muscicola for r 1295 bp, Scytonema spIAM M 262 for 1295 bp, Calothrix desertica PCC 7102 1295 bp, Calothrix spPCC 7715 r strain PCC 7715 1295 bp, Fischerella spCENA19 1295 bp, Scytonema hofmanni PCC 7110 for r 1295 bp, Fischerella PCC7414Calothrix spAHLA9 r strain AHLA9 1295 bp, Fischerella spMV9 1295 bp, Rivularia spBECID12 r strain BECID12 1295 bp, Tolypothrix spCCMP1185 for r 1295 bp, Fischerella spIAM M 263 for 1295 bp, Calothrix spXP11C r strain XP11C 1295 bp, Calothrix spBECID33 r strain BECID33 1295 bp, Calothrix spXP2B r strain XP2B 1295 bp, Calothrix spBECID16 r strain BECID16 1295 bp, Hapalosiphon sp804 1 r strain 804 1 1295 bp, Calothrix spCCMEE 5093 1295 bp, Calothrix spTJ12 UAM 372 1295 bp, Rivularia spXP27A r strain XP27A 1295 bp, Chloroglogloeopsis sp PCC7518Chlorogloeopsis spPCC 9212 for r 1295 bp, Symphyonemopsis spVAPOR1 r strain VAPOR1 1295 bp, Calothrix spBECID6 r strain BECID6 1295 bp, Filamentous thermophilic cyanobacterium tBTRCCn 101 1295 bp, Hapalosiphon welwitschii 1295 bp, Calothrix spMU27 UAM 315 1295 bp, Calothrix spXSP10A r strain XSP10A 1295 bp, Scytonema spHKAR 3 1295 bp, Calothrix spCAL3363 r strain CAL3363 1295 bp, Fischerella spHKAR 5 1295 bp, Chlorogloeopsis fritschii PCC 6912 for r 1295 bp, Mastigocladus laminosus Greenland 8 isolate 8 1295 bp, Rivularia spMU24 UAM 305 1295 bp, Nostochopsis sp89 45 r strain 89 45 1295 bp, Brasilonema spCENA114 1295 bp, Calothrix spPCC 7103 r strain PCC 7103 1295 bp, Calothrix spCAL3361 r strain CAL3361 1295 bp, Calothrix spBECID21 r strain BECID21 1295 bp, Filamentous thermophilic cyanobacterium tBTRCCn 403 1295 bp, Calothrix desertica PCC 7102 for r 1295 bp, Hapalosiphon delicatulus IAM M 266 for 1295 bp, Calothrix spXP9A r strain XP9A 1295 bp, Rivularia spPCC 7116 r strain PCC 7116 1295 bp, Fischerella major NIES 592 for 1295 bp, Stigonematales cyanobacterium AEL04 Oct 6 03 1295 bp, Calothrix spPCC 7714 r strain PCC7714 1295 bp, Brasilonema terrestre CENA116 1295 bp

Clade B1 (Chrooccidiopsis):

Unicellular thermophilic cyanobacterium tBTRCCn 23 1295 bp, Chroococcidiopsis thermalis for r 1295 bp, Chroococcidiopsis spMMG 6 1295 bp, Chroococcidiopsis spBB79 2 r SAG 2023 1295 bp, Chroococcidiopsis spCC3 complete 1295 bp, Chroococcidiopsis spBB84 1 r strain SAG 2025 1295 bp, Chroococcidiopsis spCC2 complete 1295 bp, Chroococcidiopsis spCC1 complete 1295 bp, Unicellular thermophilic cyanobacterium tBTRCCn 28 1295 bp, Chroococcidiopsis spCk4 1295 bp, Chroococcidiopsis spMMG 5 1295 bp, Chroococcidiopsis cubana r strain SAG 39 79 1295 bp, Chroococcidiopsis spBB82 3 r strain SAG 2024 1295 bp, Chroococcidiopsis spBB96 1 r strain SAG 2026 1295 bp

0.1 single species:

Phormidium autumnale UTEX 1580 1295 bp, Phormidium tergestinum CCALA 155 1295 bp, Lyngbya wollei strCarmichael/Alabama 1295 bp, Phormidium uncinatum SAG 81 79 1295 bp

Clade C1 (Arthrospira-Lyngbya):

Oscillatoria spCYA127 r strain CYA127 1295 bp, Lyngbya polychroa LP5 1295 bp, Microcoleus chthonoplastes CCY9607 1295 bp, Lyngbya majuscula HECT 1295 bp, Lyngbya spVP417a 1295 bp, Trichocoleus sociatus SAG 26 92 1295 bp, Lyngbya polychroa PNG6 68 rrnB 1295 bp, Planktothrix pseudagardhii T19 6 6 for 1295 bp, Lyngbya bouillonii PNG6 41 1295 bp, Microcoleus spDAI 1295 bp, Phormidium murrayi Ant Ph58 1295 bp, Planktothrix agardhii NIVA CYA 68 1295 bp, Microcoleus chtoenoplastes PCC7420, Microcoleus glaciei UTCC 475 1295 bp, Lyngbya sordida NAC8 51 1295 bp, Schizothrix spPNG5 22 1295 bp, Lyngbya polychroa PNG6 51 1295 bp, Planktothrix agardhii NIVA CYA 10 for 1295 bp, Planktothrix pseudagardhii HAB639 1295 bp, Symploca spHBC5 1295 bp, Planktothrix pseudagardhii T1 8 4 for 1295 bp, Symploca atlantica PCC 8002 for r 1295 bp, Microcoleus chthonoplastes EBD 1295 bp, Planktothrix rubescens r strain BC Pla 9303 1295 bp, Planktothrix mougeotii HAB3343 1295 bp, Planktothrix agardhii NIES 595 for 1295 bp, Phormidiaceae cyanobacterium CPER KK1 1295 bp, Oscillatoria spPCC 9018 1295 bp, Planktothrix rubescens PCC 10106 1295 bp, Lyngbya spVP417b 1295 bp, Microcoleus chthonoplastes CCY0002 1295 bp, Planktothrix rubescens CCAP 1459/14 1295 bp, Planktothrix pseudagardhii NIVA CYA 153 for 1295 bp, Oscillatoria spPCC 8926 1295 bp, Pseudanabaenaceae cyanobacterium ANP1 KK1 1295 bp, Lyngbya bouillonii PNG7 29 3 rrnB 1295 bp, Planktothrix pseudagardhii HAB1346 1295 bp, Lyngbya sordida NAC8 49 r 1295 bp, Planktothrix mougeotii TK4 5 for 1295 bp, Planktothrix agardhii NIVA CYA 34 for 1295 bp, Planktothrix sp2 r strain 2 1295 bp, Symploca spCCY0030 1295 bp, Planktothrix rubescens NIVA CYA 151 for 1295 bp, Lyngbya bouillonii PNG5 198 r 1295 bp, Planktothrix spPCC 9214 1295 bp, Planktothrix agardhii CCAP 1459/36 for 1295 bp, Lyngbya majuscula 3L rrnB 1295 bp, Planktothrix agardhii PCC 9637 1295 bp, Lyngbya polychroa PNG6 9 r 1295 bp, Planktothrix pseudagardhii HAB662 1295 bp, Lyngbya sordida NAC8 52 1295 bp, Lyngbya majuscula JHB 1295 bp, Planktothrix agardhii NIVA CYA 30 for 1295 bp, Microcoleus chthonoplastes CCY9606 1295 bp, Planktothrix agardhii PCC 10110 1295 bp, Planktothrix agardhii HAB326 1295 bp, Lyngbya bouillonii PNG7 63 1295 bp, Planktothrix agardhii HAB237 1295 bp, Planktothrix mougeotii TK5 1 for 1295 bp, Planktothrix pseudagardhii HAB1379 1295 bp, Lyngbya polychroa PNG6 45 1295 bp, Lyngbya polychroa PNG6 48 1295 bp, Symploca spVP642c 1295 bp, Symploca PCC 8002 1295 bp, Microcoleus chthonoplastes CCY9608 1295 bp, Planktothrix agardhii

NIVA CYA 56/3 for 1295 bp, Planktothrix pseudagardhii r isolate VR1 1295 bp, Microcoleus chthonoplastes MAF 1295 bp, Lyngbya bouillonii PAL08 16 rrnB 1295 bp, Microcoleus chthonoplastes WW11 1295 bp, Microcoleus chthonoplastes LZW 1295 bp, Lyngbya polychroa PNG5 194 rrnB 1295 bp, Oscillatoria spPCC 8954 1295 bp, Lyngbya majuscula 3L 1295 bp, Lyngbya majuscula PAB 1295 bp, Lyngbya majuscula PNG6 221 rrnB 1295 bp, Planktothrix mougeotii TR2 4 for 1295 bp, Lyngbya polychroa PNG6 68 r 1295 bp, Lyngbya polychroa PNG6 38 rrnB 1295 bp, Lyngbya bouillonii PNG7 22 r 1295 bp, Lyngbya bouillonii PNG7 34 1295 bp, Lyngbya majuscula PNG6 221 r 1295 bp, Lyngbya bouillonii PNG7 22 rrnB 1295 bp, Planktothrix agardhii CCAP 1459/23 for 1295 bp, Planktothrix pseudagardhii HAB366 1295 bp, Lyngbya majuscula NAC8 47 1295 bp, Oscillatoria spPCC 8927 1295 bp, Planktothrix spUVFP1 r strain UVFP1 1295 bp, Microcoleus chthonoplastes SAH 1295 bp, Planktothrix sp1LT27S08 r 1295 bp, Lyngbya bouillonii PNG5 198 rrnB 1295 bp, Lyngbya bouillonii PAL08 16 1295 bp, Planktothrix agardhii HAB325 1295 bp, Microcoleus chthonoplastes CCY9605 1295 bp, Microcoleus spHTT U KK5 1295 bp, Microcoleus spSAG 2212 1295 bp, Planktothrix mougeotii HAB002 1295 bp, Lyngbya polychroa PNG6 9 rrnB 1295 bp, Planktothrix pseudagardhii HAB526 1295 bp, Microcoleus chthonoplastes WW6 1295 bp, Oscillatoria spPCC 9631 1295 bp, Microcoleus chthonoplastes CCY9602 1295 bp, Lyngbya sordida NAC8 53 1295 bp, Symploca spVP642b 1295 bp, Symploca spVP642a 1295 bp, Microcoleus sp for r 1295 bp, Lyngbya bouillonii PNG7 29 3 r 1295 bp, Planktothrix agardhii 213 1295 bp, Lyngbya bouillonii PNG7 14 1295 bp, Lyngbya bouillonii PNG7 4 1295 bp, Lyngbya polychroa PNG5 192 1295 bp, Lyngbya sordida NAC8 49 rrnB 1295 bp, Planktothrix rubescens NIES 1266 1295 bp, Microcoleus chthonoplastes CCY9603 1295 bp, Microcoleus chthonoplastes WW3 1295 bp, Symploca spVP377 1295 bp, Planktothrix pseudagardhii HAB1131 1295 bp, Planktothrix pseudagardhii HAB414 1295 bp, Lyngbya polychroa PNG6 2 1295 bp, Planktothrix mougeotii HAB626 1295 bp, Lyngbya polychroa PNG6 57 1295 bp, Symploca atlantica CCY9617 1295 bp, Planktothrix agardhii NIVA CYA 313 for 1295 bp

Clade C2 (Trichodesmium-Oscillatoria):

Geitlerinema spCCY0102 1295 bp, Arthrospira platensis CG590 1295 bp, Oscillatoria miniata NAC8 50 1295 bp, Arthrospira platensis strain UTEX 2340 1295 bp, Oscillatoria spongeliae SI04 40 1295 bp, Hydrocoleum lyngbyaceum HBC7 1295 bp, Phormidium spETS 05 r isolate ETS 05 1295 bp, Oscillatoria prolifera for r 1295 bp, Arthrospira platensis strain NIES 39 1295 bp, Lyngbya hieronymusii varhieronymusii CN4 3 for 1295 bp, Lyngbya cfconfervoides VP0401 1295 bp, Oscillatoria spongeliae SI04 46 1295 bp, Planktothricoides raciborskii INBaOR for 1295 bp, Pseudoscillatoria coralii BgP10 4S 1295 bp, Geitlerinema spCCY9412 1295 bp, Oscillatoria spongeliae SI04 45 1295 bp, Phormidium autumnale Arct Ph5 1295 bp, Phormidium lumbricale UTCC 476 1295 bp, Blennothrix spPNG05 4 1295 bp, Oscillatoria margaritifera NAC8 54 1295 bp, Oscillatoria sancta for r 1295 bp, Oscillatoria spongeliae KR04 3 1295 bp, Oscillatoria spongeliae 36P3 1295 bp, Arthrospira platensis strain SAG 257 80 1295 bp, Ocfcorallinae r 1295 bp, Trichodesmium thiebautii 1295 bp, Trichodesmium sp 1295 bp, Planktothricoides raciborskii NIES 207 for 1295 bp, Oscillatoria spongeliae 32P1 1295 bp, Lyngbya aestuarii PCC 7419 for r 1295 bp, Planktothricoides raciborskii T1 6 2 for 1295 bp, Lyngbya aestuarii for r 1295 bp, Geitlerinema spBBD 1295 bp, Phormidium animale SAG 1459 6 1295 bp, Arthrospira platensis EB 9602 1295 bp, Trichodesmium erythraeum IMS101 for r 1295 bp, Trichodesmium erythraeum 1295 bp, Microcoleus vaginatus UBI KK2 1295 bp, Planktothricoides raciborskii NSLA3 for 1295 bp, Arthrospira maxima FACHB438 complete 1295 bp, Oscillatoria sancta PCC 7515 1295 bp, Geitlerinema spBBD HS217 1295 bp, Arthrospira maxima BJ 2000 1295 bp, Planktothricoides raciborskii OR1 1 for 1295 bp, Lyngbya polychroa LP16 1295 bp, Oscillatoria lutea for 1295 bp, Microcoleus vaginatus PBP D KK1 1295 bp, Microcoleus vaginatus SEV1 KK3 1295 bp, Microcoleus acremanii UTCC 313 1295 bp, Geitlerinema spBBD HS223 1295 bp, Oscillatoria spongeliae SI04 47 1295 bp, Oscillatoria acuminata for r 1295 bp, Oscillatoria spMMG 3 1295 bp, Microcoleus chthonoplastes CCY0602 1295 bp, Trichodesmium

tenue 1295 bp, Planktothricoides raciborskii NSLA4 for 1295 bp, Trichodesmium contortum 1295 bp, Tychonema bourrellyi CCAP 1459/11B for 1295 bp, Arthrospira platensis 8005Geitlerinema spA28DM 1295 bp, Phormidium autumnale SAG 35 90 1295 bp, Microcoleus antarcticus UTCC 474 1295 bp, Arthrospira jenneri EB 9604 1295 bp, Microcoleus rushforthii UTCC 296 1295 bp, Oscillatoria spCk2 1295 bp, Oscillatoria spongeliae KR04 1 1295 bp, Microcoleus vaginatus CSI U KK1 1295 bp, Oscillatoria sp195 A20 1295 bp, Phormidium spUTCC 487 1295 bp, Trichodesmium hildebrandtii 1295 bp, Oscillatoria spNAC8 18 1295 bp, Phormidium spKU003 for r 1295 bp, Arthrospira platensis strain PCC 9223 1295 bp, Phormidium autumnale SAG 78 79 1295 bp, Oscillatoria spongeliae SI04 42 1295 bp, Arthrospira platensis strain PCC 9108 1295 bp, Oscillatoria margaritifera NAC8 55 1295 bp, Oscillatoria spongeliae KR04 4 1295 bp, Oscillatoria spongeliae 310P1 1295 bp, Trichodesmium havanum strF34 5 1295 bp, Phormidium autumnale Ant Ph68 1295 bp, Geitlerinema PCC7105 for r 1295 bp, Oscillatoria spongeliae SI04 41 1295 bp, Geitlerinema spFlo1 1295 bp, Phormidium spHBC9 1295 bp, Microcoleus vaginatus SAG 2211 1295 bp, Phormidium cfterebriformis KR2003/25 1295 bp, Geitlerinema spW 1 1295 bp, Microcoleus vaginatus CJI U2 KK2 1295 bp

Clade A1 (Microcystis):

Microcystis spCHAB727 1295 bp, Microcystis ichthyoblabe for r1 1295 bp, MSU66194 Microcystis spKND9506 1295 bp, Microcystis wesenbergii for r 1295 bp, Chroococcus spJJCM 1295 bp, Microcystis spCHAB731 1295 bp, Microcystis wesenbergii NIES 107 1295 bp, Gloeothece sp PCC 6909/1 1295 bp, Microcystis spCHAB728 1295 bp, Cyanothece sp PCC 8801 1295bp, Microcystis spCHAB729 1295 bp, MAU40331 Microcystis spAWT139 1295 bp, Merismopedia glauca 0BB39S01 r strain 0BB39S01 1295 bp, MWU40334 Microcystis wesenbergii NIES112 r 1295 bp, M glauca strain B1448 1 rDNA 1295 bp, Gmembranacea r 1295 bp, Microcystis wesenbergii DNA for 1295 bp, Microcystis aeruginosa strain 038 1295 bp, Microcystis aeruginosa for strain NIES 604 1295 bp, Microcystis sp269 r strain 269 1295 bp, Microcystis viridis NIES 102 1295 bp, Cyanothece spWH 8904 1295 bp, Microcystis spCHAB1447 1295 bp, Crocosphaera watsonii WH 8501 1295 bp, AF193247 Symbiont of Climacodium frauenfeldianum 1295 bp, Microcystis novacekii for r 1295 bp, Cyanothece spPCC 7424 1295 bp, MAU40338 Microcystis aeruginosa PCC7005 r 1295 bp, Snowella litoralis 1LT47S05 r strain 1LT47S05 1295 bp, Microcystis aeruginosa NPCD 1 1295 bp, Microcystis aeruginosa 0BF29S03 r 1295 bp, Microcystis spCHAB1449 1295 bp, Snowella litoralis 0TU37S04 r strain 0TU37S04 1295 bp, Cyanothece spWH 8902 1295 bp, Microcystis aeruginosa for r isolate TAC86 1295 bp, Pleurocapsa PCC7327 for r 1295 bp, Microcystis spGL260735 1295 bp, Microcystis ichthyoblabe 0BB35S01 r 1295 bp, Microcystis spCHAB657 1295 bp, Microcystis novacekii for r isolate TAC65 1295 bp, Cyanothece sp r strain PCC 7424 1295 bp, MAU03402 Microcystis aeruginosa PCC7806 r 1295 bp, Microcystis aeruginosa 0BB35S02 r 1295 bp, Synechocystis PCC6805 for r 1295 bp, MAU40339 Microcystis aeruginosa PCC7820 r 1295 bp, Microcystis novacekii for r isolate TAC20 1295 bp, Microcystis sp for r isolate 4B3 1295 bp, Microcystis aeruginosa strain SPC 777 1295 bp, Microcystis sp for r isolate T17 1 1295 bp, Microcystis aeruginosa r isolate MK10 10 1295 bp, Microcystis aeruginosa DNA for 1295 bp, Microcystis spCHAB726 1295 bp, Microcystis viridis DNA for 1295 bp, Microcystis sp130 r strain 130 1295 bp, Microcystis aeruginosa for 1295 bp, Synechocystis spCCALA 700 1295 bp, Snowella rosea 1LM40S01 r strain 1LM40S01 1295 bp, Woronichinia naegeliana 0LE35S01 r strain 0LE35S01 1295 bp, Microcystis ichthyoblabe for r 1295 bp, Aphanothece sacrum for r 1295 bp, Synechocystis PCC6803 MAU03403 Microcystis aeruginosa NIES89 r 1295 bp, MVU40332 Microcystis viridis NIES 102 complete 1295 bp, Radiocystis spJJ30 1295 bp, MWU40333 Microcystis wesenbergii NIES 107 complete 1295 bp, MAU40340 Microcystis aeruginosa PCC 7941 complete 1295 bp, Microcystis sp for r isolate T1 4 1295 bp, Microcystis spGL280641 r strain GL280641 1295 bp, Aphanocapsa spHBC6 1295 bp

Clade C3 (Spirulina):

Spirulina spstrain MPI S4 r 1295 bp, Spirulina major 0BB36S18 r strain 0BB36S18 1295 bp, Spirulina subsalsa IAM M 223 for 1295 bp, Spirulina PCC6313, Spirulina spGLS010 1295 bp, Halospirulina sp CCC Baja 95 Cl 3 r 1295 bp, Spirulina spstrain CCC Snake PY 85 r 1295 bp, Halospirulina tapeticola CCC Baja 95 Cl 2 1295 bp, Halospirulina sp MPI S3 r 1295 bp

Clade A2 (Synechocystis):

Synechocystis PCC6308 1295 bp, Synechococcus spHOG 1295 bp, Cyanobacterium spMBIC10216 for r 1295 bp, Synechococcus spPCC 8807 1295 bp, Cyanothece sp115 1295 bp, Synechococcus spPH40 1295 bp, Synechococcus sp for r strain PCC73109 1295 bp, Halothece spPCyano42 1295 bp, Cyanothece sp r strain PCC 7418 1295 bp, Synechococcus elongatus CCMP1630 1295 bp, Synechococcus spT71 1295 bp, Synechococcus spUH7 1295 bp, Euhalothece spZ M001 1295 bp, Dactylococcopsis sp PCC 8305 1295 bp, Euhalothece sp r strain MPI 96N304 1295 bp, Cyanothece sp109 1295 bp, Rubidibacter lacunae KORDI 51 2 1295 bp, Oscillatoria rosea IAM M 220 for 1295 bp, Cyanothece spGSL007 1295 bp, Synechocystis trididemni for r 1295 bp, Euhalothece sp r strain MPI 95AH10 1295 bp, Cyanothece sp113 1295 bp, Phormidium spMBIC10210 for r strain MBIC10210 1295 bp, Halothece spPCC 7418 1295 bp, Prochloron sp for 1295 bp, Cyanothece sp104 1295 bp, Synechococcus sp for r strain PCC7003 1295 bp, Synechococcus sp r strain PCC 7002 1295 bp, Euhalothece sp r strain MPI 95AH13 1295 bp, Synechococcus sp for r strain PCC7117 1295 bp, Halothece sp r strain MPI 96P605 1295 bp

Clade B2 (Pleurocapsa):

Pleurocapsa minor r strain SAG 4 99 1295 bp, Dermocarpella incrassata r strain SAG 29 84 1295 bp, Dermocarpa spMBIC10768 for r 1295 bp, Chroococcidiopsis spPCC 6712 r strain PCC 6712 1295 bp, Myxosarcina PCC 7325 1295 bp, Chroococcidiopsis PCC6712 for r 1295 bp, Stanieria PCC7301 for r 1295 bp, Myxosarcina PCC 7312 1295 bp, Dermocarpa spMBIC10004 for r 1295 bp, Pleurocapsa spCALU 1126 1295 bp, Pleurocapsa spOU 12 1295 bp, Pleurocapsa PCC7319 for r 1295 bp, Pleurocapsa sp r 1295 bp, Pleurocapsa spOU 11 1295 bp, Chroococcidiopsis spCCMP1489 r strain CCMP1489 1295 bp

single species:

Xenococcus spCR 15M 1295 bp, Stanieria cyanosphaera PCC 7437 1295 bp, Stanieria cyanosphaera for r 1295 bp

Clade A3 (Chroococcus):

Chroococcus spJJCV r strain JJCV 1295 bp, Chroococcus spCCALA 702 1295 bp, Chroococcus spCCALA 701 1295 bp, Chroococcus spCCALA 057 1295 bp, Chroococcus cfmembraninus CCALA 054 1295 bp, Gloeocapsa PCC73106 for r 1295 bp, Limnococcus limneticus Svet06 1295 bp, Chroococcus minutus CCALA

055 1295 bp

Clade A4 (Synechococcus):

Cyanobacterium 5X15 r 1295 bp, Synechococcus spKORDI 70 1295 bp, Prochlorococcus marinus subsppastoris stTL2 1295 bp, Synechococcus spKORDI 15 1295 bp, Synechococcus spKORDI 49 1295 bp, Synechococcus sp for r strain PCC7001 1295 bp, AF098372 Synechococcus like strABRAXAS 1295 bp, Cyanobium spJJ2 3 r strain JJ2 3 1295 bp, Cyanobium sp0BB42S04 r strain 0BB42S04 1295 bp, MHU40336 Microcystis holsatica NIES43 r 1295 bp, Cyanobium spJJ9 A3 r strain JJ9 A3 1295 bp, Synechococcus spKORDI 63 1295 bp, Synechococcus PCC7943 1295 bp, Synechococcus spBN39 1295 bp, AF098370 Synechococcus like strACE 1295 bp, Synechococcus spHOS 1295 bp, AF115268 Prochlorococcus spMIT9201 1295 bp, Synechococcus spTAGS 1295 bp, AF001467 Prochlorococcus spNATL2A 1295 bp, Prochlorococcus marinus strTAK9803 2 1295 bp, Syn CC9605Synechococcus spCENA108 1295 bp, Synechococcus spKORDI 18 1295 bp, Syn WH8101Synechococcus spKORDI 11 1295 bp, AF098374 Synechococcus like strP212 1295 bp, Synechococcus spKORDI 12 1295 bp, Cyanobium spJJ27STR r strain JJ27STR 1295 bp, AF001470 Prochlorococcus spNATL2 1295 bp, Synechococcus spRS9921 1295 bp, Synechococcus spKORDI 13 1295 bp, Cyanobium spJJ9 C6 r strain JJ9 C6 1295 bp, Cyanobium spJJ17 5 r strain JJ17 5 1295 bp, AF081834 Synechococcus WH7803 1295 bp, AF001474 Prochlorococcus spTATL1A 1295 bp, Cyanobium spJJ5 5 r strain JJ5 5 1295 bp, Synechococcus spOli31 1295 bp, Synechococcus spMinos11 1295 bp, Synechococcus spPS838 1295 bp, Cyanobium spJJM10A4 r strain JJM10A4 1295 bp, Cyanobium spY0011 1295 bp, Synechococcus sp1tu39s01 r strain 1tu39s01 1295 bp, Synechococcus spBN31 1295 bp, Cyanobium spJJ12A2 r strain JJ12A2 1295 bp, Synechococcus spMBIC10459 for r 1295 bp, Cyanobium spJJ19B5 r strain JJ19B5 1295 bp, Synechococcus spKORDI 54 1295 bp, Synechococcus sp0tu28s07 r strain 0tu28s07 1295 bp, Synechococcus spKORDI 16 1295 bp, Cyanobium spJJ15 13 r strain JJ15 13 1295 bp, AF115270 Prochlorococcus spMIT9211 1295 bp, P marinus for 1295 bp, Synechococcus spEW15 1295 bp, MEU40335 Microcystis elabens NIES42 r 1295 bp, AF001479 Synechococcus WH8103 1295 bp, Synechococcus spBAC 9803 1295 bp, Synechococcus spBAC 9810 1295 bp, Synechococcus spKORDI 71 1295 bp, Cyanobium spJJM10D5 r strain JJM10D5 1295 bp, Synechococcus spKORDI 30 1295 bp, Cyanobium spJJ11D3 r strain JJ11D3 1295 bp, Aphanothece sp0BB21S01 r strain 0BB21S01 1295 bp, Synechococcus spPS723 1295 bp, Synechococcus spMBIC10007 for r 1295 bp, Synechococcus spMBIC10456 for r 1295 bp, Synechococcus spBE0807I 1295 bp, Synechococcus PCC9005 1295 bp, Synechococcus PCC7009 1295 bp, Synechococcus spWH 5701 1295 bp, Synechococcus spRS9918 1295 bp, Synechococcus spMBIC10089 for r 1295 bp, AF098373 Synechococcus like strP211 1295 bp, AF053397 Prochlorococcus spMIT9303 1295 bp, Synechococcus sp0BB26S03 r strain 0BB26S03 1295 bp, AF001478 Synechococcus WH7805 1295 bp, Synechococcus spPS715 1295 bp, Synechococcus sp0BB22S0 r strain 0BB22S05 1295 bp, Synechococcus spC129 1295 bp, Microcystis elabens DNA for 1295 bp, Synechococcus spMLCB 1295 bp, Cyanobium spLB03 1295 bp, Microcystis holsatica DNA for 1295 bp, AF098371 Synechococcus like strPENDANT 1295 bp, Synechococcus spWH 8103 1295 bp, Synechococcus spWH 8018 1295 bp, Synechococcus spKORDI 52 1295 bp, Synechococcus PCC7918 1295 bp, AF001468 Prochlorococcus spMIT9107 1295 bp, AF001472 Prochlorococcus spGP2 1295 bp, Synechococcus spWH 7803 1295 bp, Prochlorococcus marinus strEQPAC1 1295 bp, Prochlorococcus spNATL1 1295 bp, Cyanobium spJJ12 6 r strain JJ12 6 1295 bp, Synechococcus spPS680 1295 bp, Synechococcus spAlmo3 1295 bp, Cyanobium spJJ7 9 r strain JJ7 9 1295 bp, Synechococcus spKORDI 17 1295 bp, Synechococcus spNAN 1295 bp, Cyanobium spNS01 1295 bp, Synechococcus spMBIC10613 for r strain MBIC10613 1295 bp, Cyanobium spJJ22 10 r strain JJ22 10 1295 bp, Synechococcus spRS9920 1295 bp, Synechococcus spKORDI 56 1295 bp, Synechococcus spMinos12 1295 bp, AF001475 Prochlorococcus spTATL1B 1295 bp, Synechococcus spIOCAS0401 1295 bp, Prochlorococcus marinus subsppastoris stTL1 1295 bp, Synechococcus spTAG 1295 bp, Synechococcus spWH 8020 1295 bp, Synechococcus spUBR 1295 bp, Synechococcus

spKORDI 36 1295 bp, Synechococcus spKUAC 3043 1295 bp, Cyanobium spPCC 6904 1295 bp, Cyanobium spJJ13 6 r strain JJ13 6 1295 bp, AF115269 Prochlorococcus spMIT9202 1295 bp, AF053396 Prochlorococcus spMIT9302 1295 bp, Synechococcus spACT9807 1295 bp, Prochlorococcus spCCMP1426 1295 bp, Synechococcus spWH 8016 1295 bp, Synechococcus spKORDI 50 1295 bp, Synechococcus spCCMP839 1295 bp, Cyanobium spGSL004 1295 bp, Syn PCC6301 Cyanobium spJJ10 3 r strain JJ10 3 1295 bp, Synechococcus spT7cc1 1295 bp, AF053398 Prochlorococcus spMIT9312 1295 bp, Cyanobium spJJ19 11 r strain JJ19 11 1295 bp, Synechococcus spGL150636 r strain GL150636 1295 bp, Cyanobium spJJNV r strain JJNV 1295 bp, AF053399 Prochlorococcus spMIT9313 1295 bp, AF115271 Prochlorococcus spMIT9215 1295 bp, Synechococcus spPS845 1295 bp, Cyanobium spJJM10D4 r strain JJM10D4 1295 bp, Synechococcus spKORDI 65 1295 bp, Synechococcus spKORDI 78 1295 bp, Synechococcus spKORDI 42 1295 bp, Cyanobium spJJ22K r strain JJ22K 1295 bp, Merismopedia spCENA106 1295 bp, Synechococcus spPS840 1295 bp, Synechococcus spIOCAS0402 1295 bp, Cyanodictyon spJJCD r strain JJCD 1295 bp, AF001471 Prochlorococcus spPAC1 1295 bp, Synechococcus PCC7920 1295 bp, Cyanobium spJJ9 5 r strain JJ9 5 1295 bp, Cyanobium spJJR2A5 r strain JJR2A5 1295 bp, Synechococcus spWH 8012 1295 bp, Synechococcus spPCC 8966 1295 bp, Cyanobium spJJ8 4 r strain JJ8 4 1295 bp

Clade AC1 :

Leptolyngbya sp 1295 bp, Synechococcus spMBIC10598 for r strain MBIC10598 1295 bp, MSU66194 Microcystis spKND9506 1295 bp, Microcystis wesenbergii for r 1295 bp, Leptolyngbya spCENA112 1295 bp, FilthermocyanobacteriumLeptolyngbya tenerrima UTCC 77 1295 bp, Leptolyngbya spCENA103 1295 bp, CfLeptolyngbya spGreenland 9 1295 bp, Limnothrix spB15 1295 bp, Leptolyngbya spFYG 1295 bp, MWU40334 Microcystis wesenbergii NIES112 r 1295 bp, Filamentous thermophilic cyanobacterium tBTRCCn 102 1295 bp, Calothrix spKVSF5 1295 bp, Arthronema africanum SAG 12 89 for 1295 bp, Limnothrix redekei 165c r strain 165c 1295 bp, Leptolyngbya spHBC3 1295 bp, L foveolarum r 1295 bp, AF170757 LPP group cyanobacterium QSSC5cya 1295 bp, Phormidium spIAM M 99 for 1295 bp, Leptolyngbya angustata UTCC 473 1295 bp, Oscillatoriales cyanobacterium OU 4 1295 bp, Leptolyngbya ant lh52 1Leptolyngbya spOU 14 1295 bp, Leptolyngbya spBN34 1295 bp, Phormidium spOU 10 1295 bp, Oscillatoriales cyanobacterium UVFP2 r strain UVFP2 1295 bp, Microcystis wesenbergii DNA for 1295 bp, Limnothrix spCENA110 1295 bp, Leptolyngbya spOU 8 1295 bp, Filamentous cyanobacterium GSL035 1295 bp, Microcystis sp269 r strain 269 1295 bp, Halomicronema spSCyano39 1295 bp, Starria zimbabweensis SAG 74 90 for r 1295 bp, Limnothrix spCENA109 1295 bp, Pseudanabaena tremula UTCC 471 1295 bp, Synechococcus spIR11 1295 bp, Chamaesiphon PCC7430Leptolyngbya spBN44 1295 bp, Phormidium spMBIC10818 for r strain MBIC10818 1295 bp, PlectonemMAU03402 Microcystis aeruginosa PCC7806 r 1295 bp, Leptolyngbya spNi6 C2 for r 1295 bp, CfLeptolyngbya spGreenland 10 1295 bp, Leptolyngbya spGreenland 7 1295 bp, Halomicronema spTFEP1 1295 bp, Leptolyngbya spBN43 1295 bp, Synechococcus C9Pseudanabaenaceae cyanobacterium PBP U KK1 1295 bp, Microcystis aeruginosa DNA for 1295 bp, L boryanum 1295 bp, Oscillatoriales cyanobacterium JSC 1 1295 bp, Filamentous thermophilic cyanobacterium tBTRCCn 408 1295 bp, Synechococcus sp for r strain PCC7335 1295 bp, Filamentous thermophilic cyanobacterium tBTRCCn 302 1295 bp, Leptolyngbya spOU 13 1295 bp, LPP group MBIC10086 for r 1295 bp, Oscillatoria spIAM M 117 for 1295 bp, Phormidium spMBIC10002 for r 1295 bp, Limnothrix redekei 165a r strain 165a 1295 bp, AF170758 LPP group cyanobacterium QSSC8cya 1295 bp, Phormidium spSAG 37 90 1295 bp, Leptolyngbya spMMG 1 1295 bp, Leptolyngbya spCENA104 1295 bp, Planktothrix sp FP1 1295 bp, LPP group MBIC10597 for r 1295 bp, AF170756 LPP group cyanobacterium QSSC3cya 1295 bp, Crinalium magnum SAG 34 87 for 1295 bp, MAU40340 Microcystis aeruginosa PCC 7941 complete 1295 bp, Leptolyngbya spITAC101 1295 bp, Filamentous thermophilic cyanobacterium tBTRCCn 407 1295 bp

Clade A5 (Acaryochloris):

Acaryochloris spAwaji 1 for r 1295 bp, Acaryochloris marina strain MBIC 11017 1295 bp, Synechococcus spPCC 6312 1295 bp, Acaryochloris spJJ7 5 r strain JJ7 5 1295 bp, Syn lividusC1, Thermosynechococcus BP 1, Acaryochloris marina for r strain MBIC11017 1295 bp, Synechococcus PCC6717 1295 bp, Acaryochloris spJJ8A6 r strain JJ8A6 1295 bp

Clade C4 (Pseudanabaena):

Anabaena spiroides NIES 78 1295 bp, Filamentous cyanobacterium 72 1 1295 bp, Pseudanabaena PCC7367 for r 1295 bp, Synechococcus spPCC 7502 1295 bp, Pseudanabaena PCC7403 for r 1295 bp, Arthronema gygaxiana UTCC 393 1295 bp, Pseudanabaena sp63 1 1295 bp, Oscillatoria limnetica strain MR1 r 1295 bp, Pseudanabaena sp1tu24s9 r strain 1tu24s9 1295 bp, Phormidium mucicola IAM M 221 1295 bp, Pseudanabaena PCC6802 1295 bp, Pseudanabaena PCC7408 for r 1295 bp, Limnothrix redekei NIVA CYA 227/1 for 1295 bp, Pseudanabaena spPCC 7403 1295 bp, Pseudanabaena sp0tu30s18 r strain 0tu30s18 1295 bp, Pseudanabaena PCC6903 for r 1295 bp

Chloroplasts:

Oedogonium cardiacum chloroplast , Prochlorales cyanobacterium EV 7 1295 bp, Glaucocystis nostochinearum plastid 16S rRNA gene, Parachlorella kessleri chloroplast , Stigeoclonium helveticum strain UTEX 441 chloroplast , Porphyra yezoensis chloroplast

Single species at the basis:

Gloeobacter violaceus PCC 7421, Synechococcus sp. PCC 7336 1295 bp, Synechococcus sp. P1

Non-cyanobacteria:

Chlorobium sp., Agrobacterium tumefaciens, Beggiatoa sp. 'Chiprana', Candidatus Chlorothrix halophila, Thiobacillus prosperus, Escherichia coli HS

Additional File 2.2 - Phylogenetic tree of cyanobacteria - newick format

Phylogenetic tree of 1,254 cyanobacterial sequences including six chloroplasts and six Eubacteria analyzed using maximum likelihood analysis with a GTR+G+I estimated substitution model, conducted with the software RAxML. (The file is to be found under the following link: *http://www.biomedcentral.com/1471-2148/11/45/additional*)

Additional File 2.3 - Rooted Bayesian consensus tree of 27 eubacterial species including five cyanobacterial species

Tree was reconstructed from 16S rRNA gene sequences based on GTR+I+G substitution model with an archaean outgroup. Posterior probabilities (black) and bootstrap values (red) from 100 re-samplings are displayed at the nodes. Cyanobacteria (blue-green box) are strongly supported as a monophyletic group with *Gloeobacter violaceus* being closest to other eubacterial species.

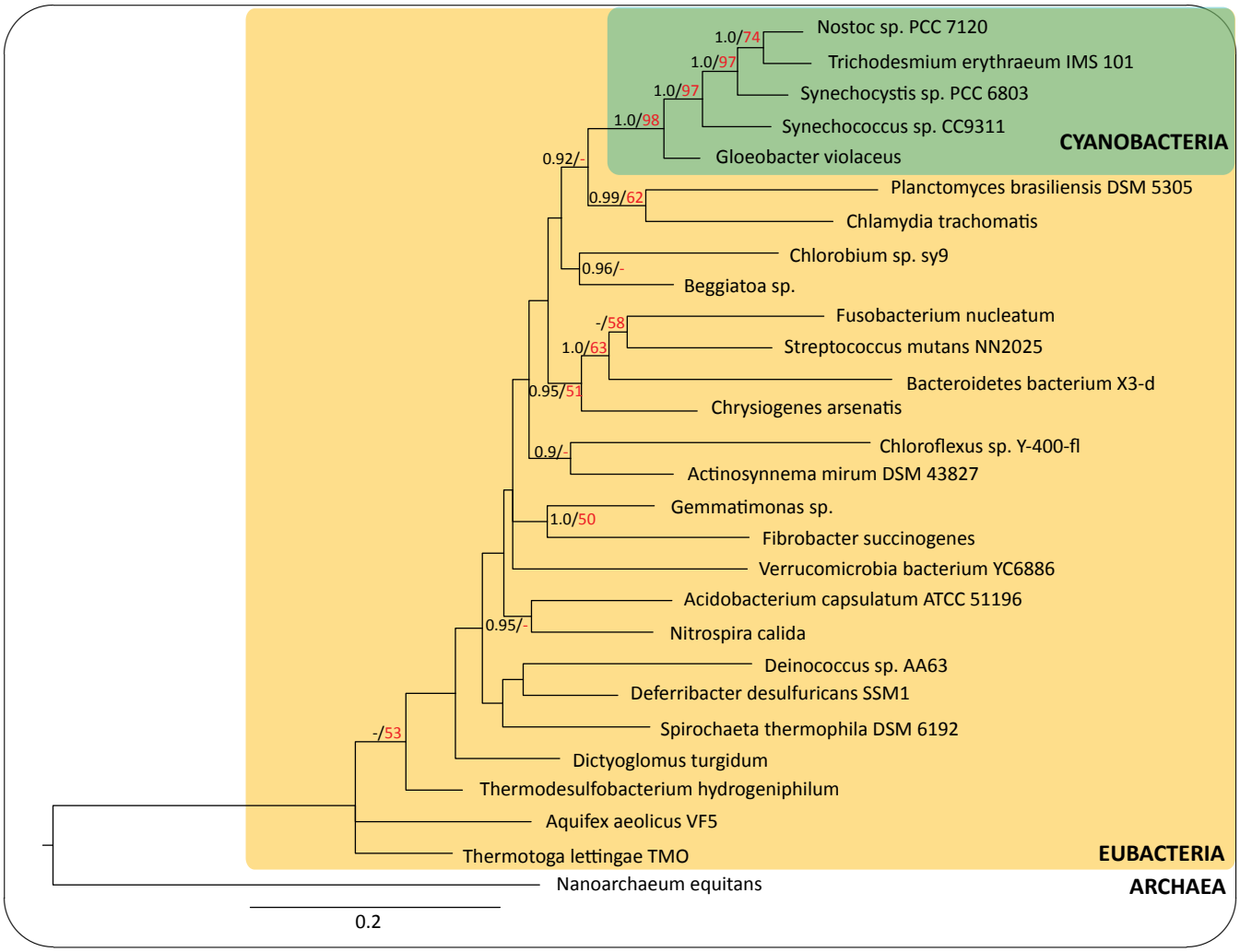


Fig. A.8. Rooted Bayesian consensus tree of 27 eubacterial species including five cyanobacterial species.

Additional File 2.4 - Bayesian consensus trees of cyanobacterial subset and different outgroups - newick format

22 Bayesian consensus trees with posterior probabilities of a cyanobacterial subset (58 taxa) and different eubacterial outgroups, displayed in newick format. Trees were run for 10,000,000 generations using a GTR+I+G substitution model with the first 3,000,000 generations being discarded as a burn-in. (The file is to be found under the following link: *http://www.biomedcentral.com/1471-2148/11/45/additional*)

Additional File 2.5 - Results from the test of substitutional saturation

Substitutional saturation of the sequences was tested using DAMBE software. The index of substitutional saturation is smaller than the estimated critical value irrespective of the symmetry of the tree. The sequences are therefore not saturated.

Symmetrical tree	
Proportion of invariable sites	0.038
mean H[1]	0.33
Standard error	0.018
ISS[2]	0.173
ISSc[3]	0.756
Asymmetrical tree	
ISSc[3]	0.460

[a]Mean entropy for all sites
[b]entropy based index of substitutional saturation
[c]critical value for the index of substitutional saturation

Additional File 2.6 - Maximum likelihood tree of cyanobacterial subset

Maximum likelihood analysis of 16S rDNA sequences from 58 cyanobacteria, based on GTR+G+I substitution model, with *Beggiatoa sp.* as an outgroup. Posterior probabilities (> 0.9) in black and bootstrap values ($> 50\%$) in red are shown at the nodes. Posterior probabilities were calculated from 265,858 trees inferred by Bayesian analysis. Bootstrap values were calculated from 500 re-samplings of the data set. Colors define groups: yellow are single-celled cyanobacteria of section I; orange single-celled from section II; green are multicellular, undifferentiated cyanobacteria from section III; blue are multicellular and differentiated bacteria from section IV; and pink from section V. Sections as described by Castenholz 2001 [64]. AC, B, C, E and E1 denote clades discussed in the text.

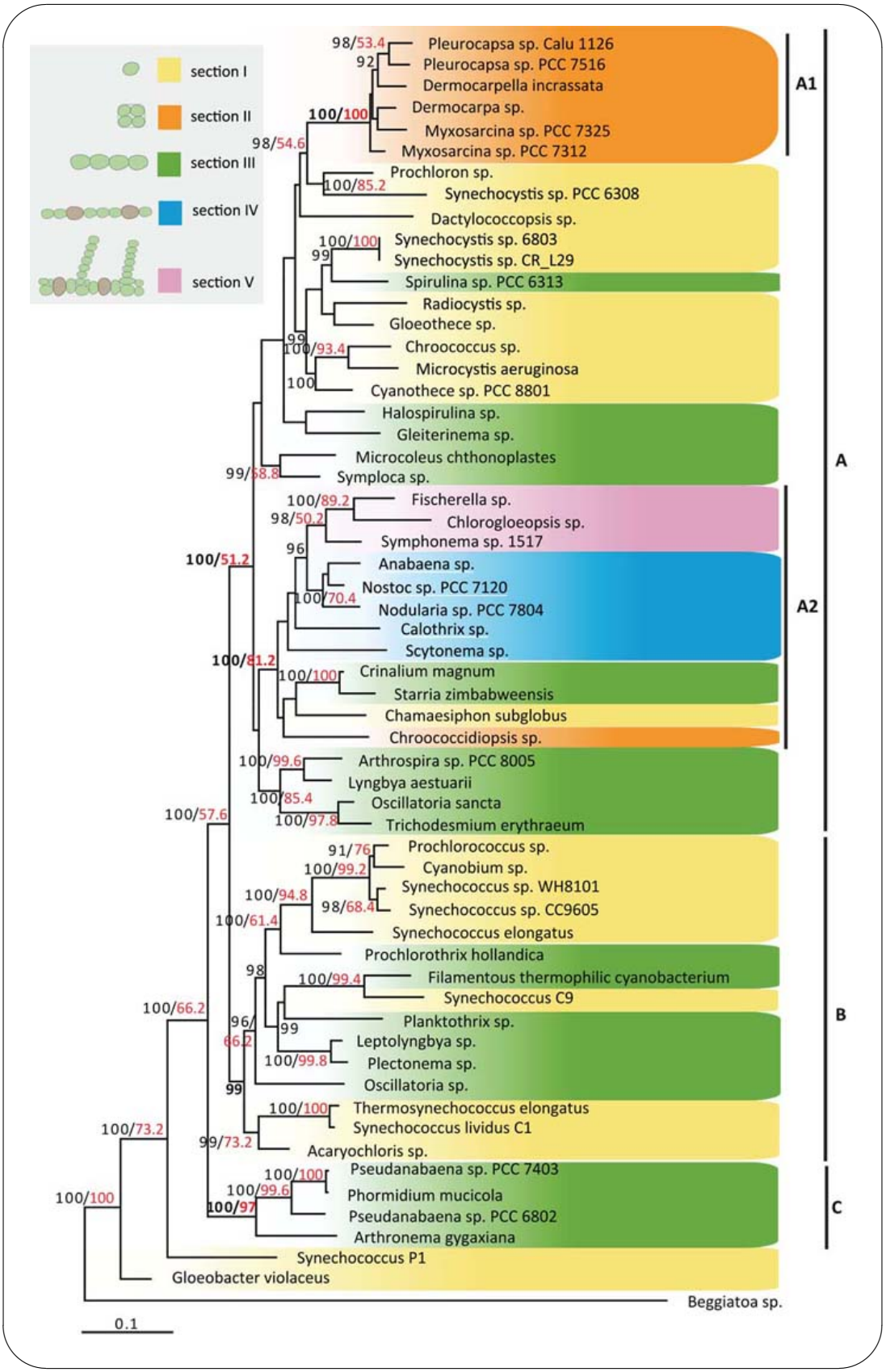


Fig. A.9. Maximum Likelihood tree of cyanobacterial subset.

Additional File 2.7 - Ancestral character state reconstruction using maximum parsimony

Maximum likelihood analysis of 16S rDNA sequences from 58 cyanobacteria, based on GTR+G+I substitution model, with *Beggiatoa sp.* as an outgroup. Posterior probabilities (> 0.9) in black and bootstrap values ($> 50\%$) in red are shown at the nodes. Posterior probabilities were calculated from 265,858 trees inferred by Bayesian analysis. Bootstrap values were calculated from 500 re-samplings of the data set. Colors define groups: yellow are single-celled cyanobacteria of section I; orange single-celled from section II; green are multicellular, undifferentiated cyanobacteria from section III; blue are multicellular and differentiated bacteria from section IV; and pink from section V. Sections as described by Castenholz 2001 [64]. AC, B, C, E and E1 denote clades discussed in the text.

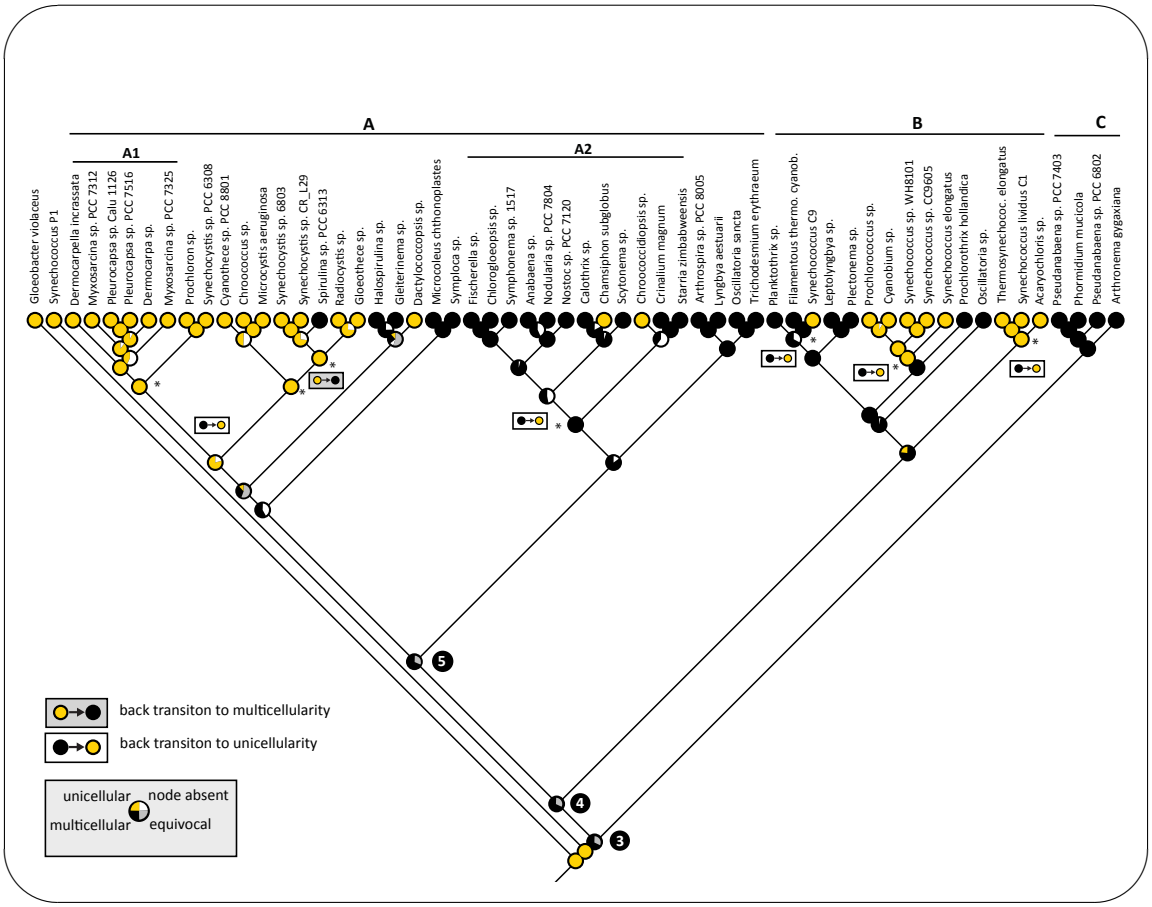


Fig. A.10. Ancestral character state reconstruction using maximum parsimony

Additional Files to chapter IV: Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event

Additional File 4.1 - Bayesian consensus tree.

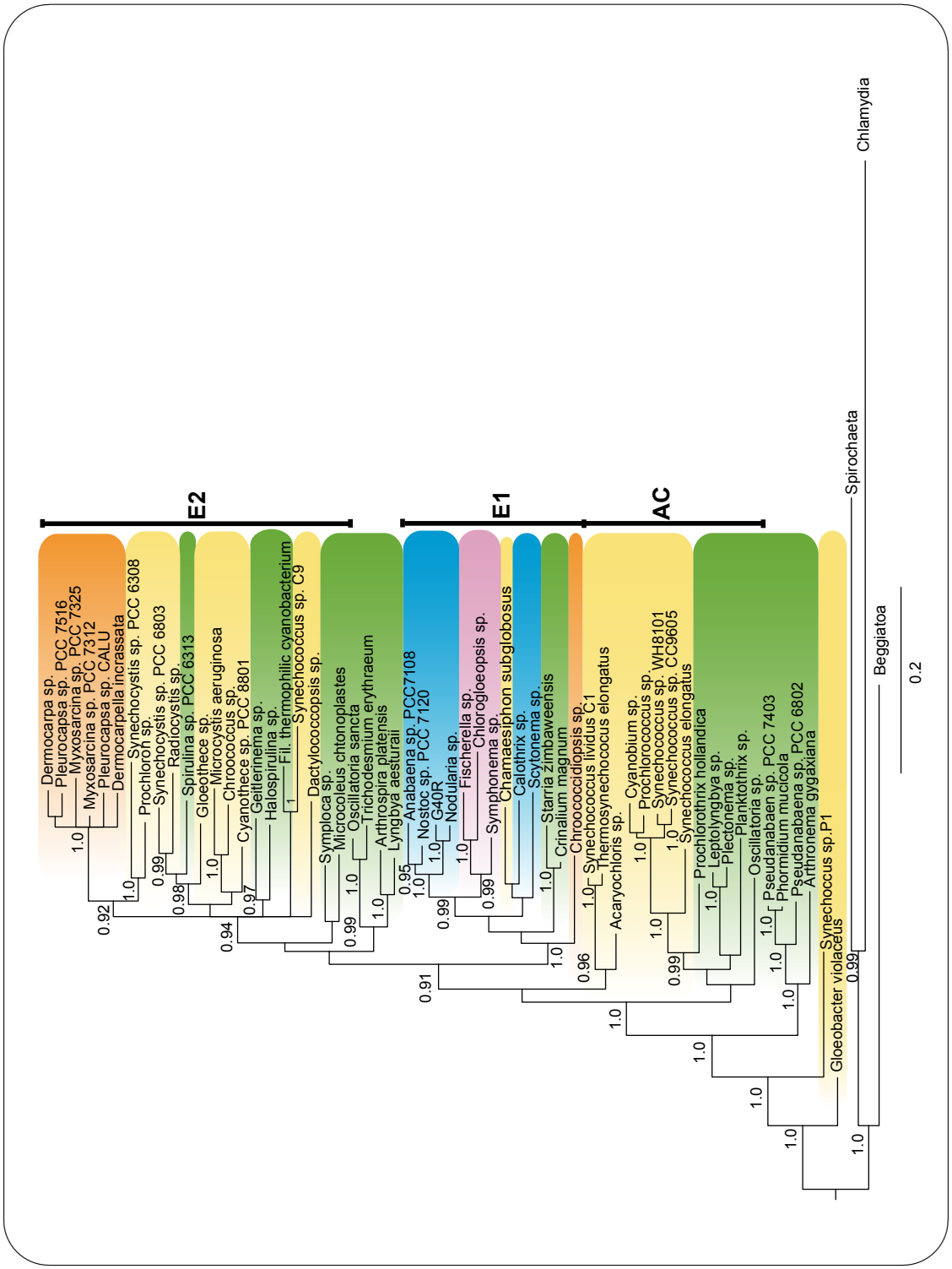


Fig. A.11. Bayesian consensus tree. Posterior probabilities > 0.9 are presented at the nodes.

Additional File 4.2 - Bayesian consensus tree of BEAST-analysis 7

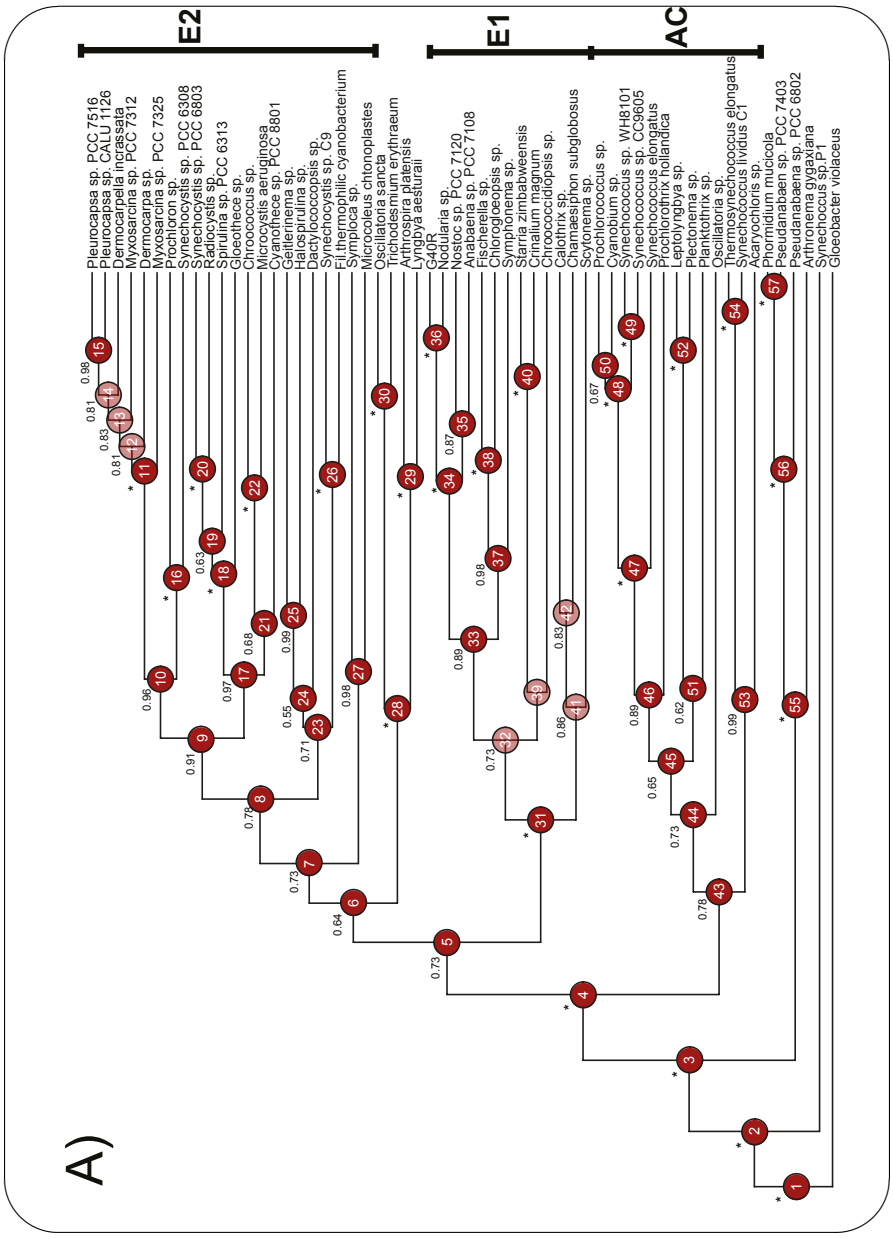


Fig. A.12. Bayesian consensus tree of BEAST-analysis 7. Posterior probability and node numbers are presented at nodes. Transparent node numbers were not recovered by all analyses.

Additional File 4.3 - Estimated clade specific diversification rates using species numbers.

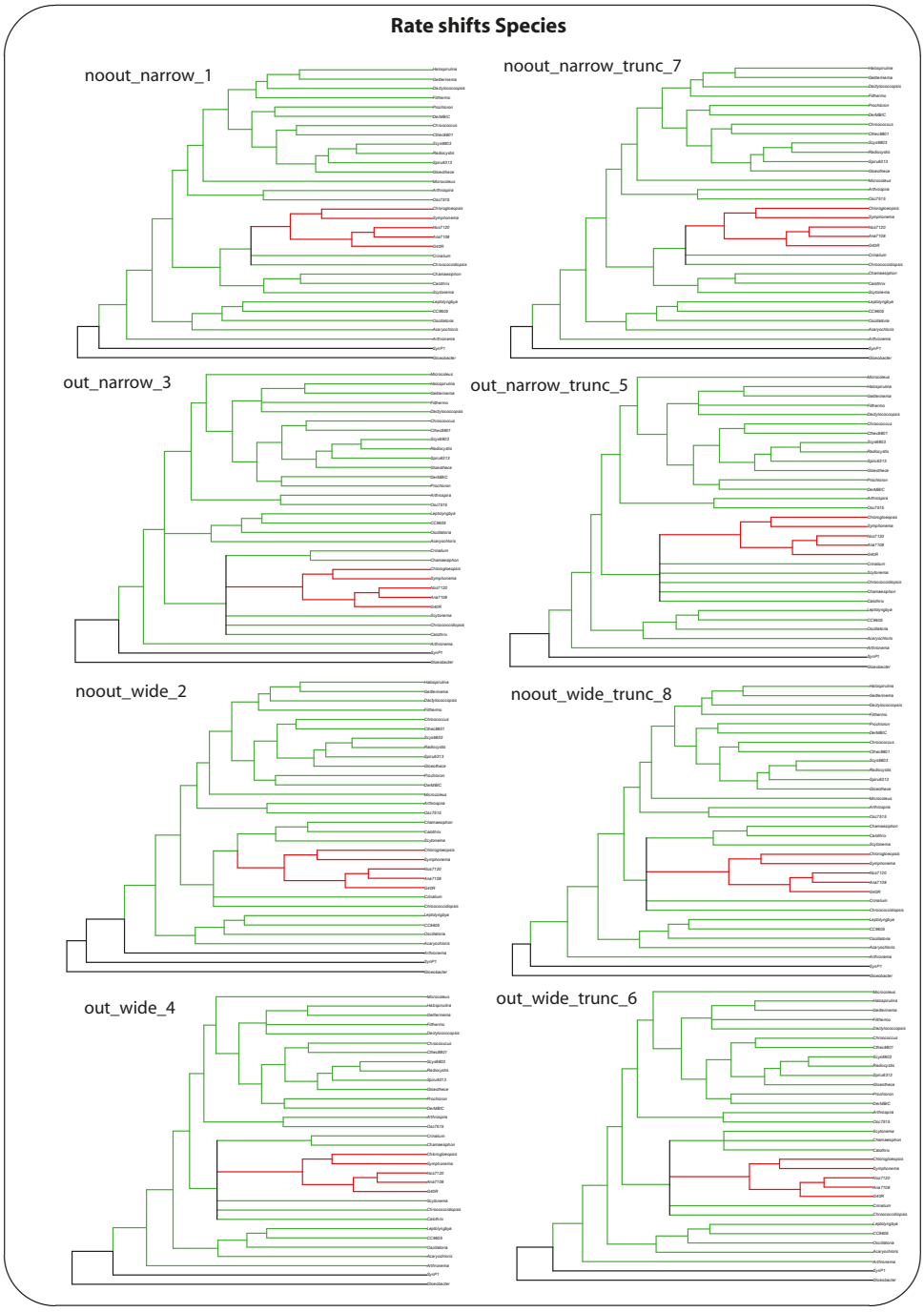


Fig. A.13. Estimated clade specific diversification rates using species numbers. Results from the diversification rate shifts for the different Bayesian consensus from the analyses.

Additional File 4.4 - Estimated clade specific diversification rates using strain numbers.

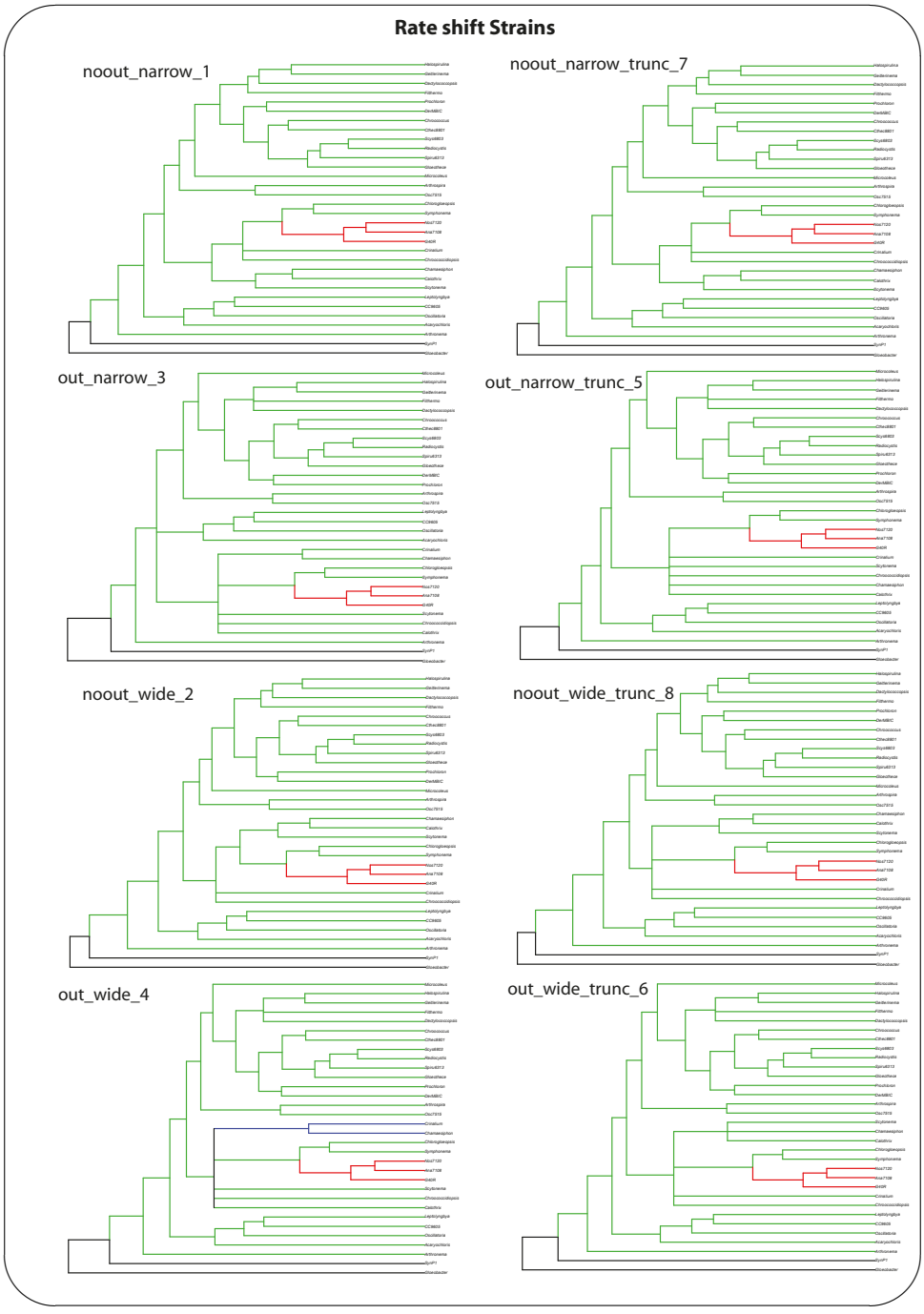


Fig. A.14. Estimated clade specific diversification rates using strain numbers. Results from the diversification rate shifts for the different Bayesian consensus from the analyses.

Additional File 4.5 - Estimated ages of nodes found in the Bayesian consensus trees for each analysis.

	1			2			3			4			5			6			7			8		
Nd	m̄	lo	up	m̄	lo	up	m̄	lo	up	m̄	lo	up	m̄	lo	up	m̄	lo	up	m̄	lo	up	m̄	lo	up
1	2.95	2.5	3.6	3.67	2.79	4.74	2.99	2.57	3.55	3.35	2.74	4.15	2.87	2.53	3.30	3.06	2.66	3.53	2.95	2.53	3.55	3.39	2.87	3.80
2	2.77	2.42	3.29	3.47	2.67	4.40	2.63	2.35	2.98	2.96	2.49	3.61	2.56	2.33	2.84	2.75	2.44	3.14	2.77	2.43	3.28	3.22	2.72	3.72
3	2.54	2.28	2.98	3.08	2.42	3.84	2.42	2.21	2.73	2.65	2.28	3.18	2.38	2.20	2.62	2.49	2.26	2.81	2.54	2.29	2.97	2.86	2.43	3.34
4	2.33	2.14	2.7	2.76	2.21	3.39	2.24	2.12	2.47	2.40	2.14	2.84	2.22	2.12	2.39	2.28	2.13	2.54	2.33	2.14	2.68	2.58	2.21	3.01
5	2.16	2.1	2.45	2.50	2.10	3.02				2.24	2.10	2.60	2.14	2.10	2.25	2.16	2.10	2.37	2.16	2.10	2.44	2.33	2.10	2.70
6	2.04	1.77	2.35	2.33	1.89	2.87	2.02	1.72	2.28	2.10	1.78	2.54	1.99	1.67	2.22	2.02	1.70	2.32	2.04	1.79	2.35	2.18	1.86	2.60
7	1.91	1.62	2.25	2.21	1.74	2.78	1.89	1.57	2.17	1.99	1.63	2.41	1.85	1.53	2.13	1.89	1.56	2.21	1.91	1.62	2.24	2.07	1.71	2.50
8	1.7	1.41	2.03	1.98	1.53	2.53	1.67	1.35	1.99	1.77	1.41	2.20	1.61	1.29	1.92	1.65	1.31	1.99	1.70	1.41	2.03	1.85	1.51	2.26
9	1.5	1.2	1.82	1.75	1.32	2.26	1.46	1.14	1.79	1.56	1.19	1.97	1.40	1.08	1.72	1.43	1.09	1.76	1.50	1.20	1.82	1.64	1.29	2.03
10	1.31	1	1.66	1.53	1.09	2.02	1.26	0.91	1.62	1.35	0.95	1.76	1.19	0.85	1.54	1.22	0.87	1.59	1.31	0.99	1.65	1.44	1.08	1.83
11	0.64	0.43	0.88	0.75	0.48	1.07	0.58	0.38	0.84	0.63	0.39	0.91	0.56	0.34	0.81	0.57	0.36	0.83	0.64	0.43	0.88	0.70	0.47	0.98
12	0.56	0.37	0.78	0.66	0.42	0.94	-	-	-	-	-	-	-	-	-	-	-	-	0.56	0.38	0.78	0.62	0.40	0.86
13	0.48	0.31	0.67	0.56	0.34	0.81	-	-	-	-	-	-	-	-	-	-	-	-	0.47	0.31	0.67	0.52	0.33	0.74
14	0.39	0.24	0.58	0.46	0.27	0.70	-	-	-	-	-	-	-	-	-	-	-	-	0.39	0.24	0.57	0.43	0.26	0.64
15	0.25	0.13	0.41	0.29	0.15	0.49	0.26	0.12	0.45	0.28	0.12	0.48	0.24	0.10	0.43	0.25	0.11	0.44	0.25	0.13	0.40	0.27	0.14	0.44
16	0.98	0.62	1.35	1.14	0.70	1.65	0.91	0.52	1.30	0.98	0.55	1.43	0.85	0.45	1.24	0.87	0.47	1.28	0.98	0.61	1.34	1.08	0.68	1.50
17	1.3	0.99	1.62	1.51	1.10	1.99	1.25	0.93	1.58	1.34	0.98	1.74	1.19	0.87	1.52	1.22	0.89	1.56	1.29	0.99	1.61	1.42	1.08	1.80
18	0.97	0.68	1.3	1.13	0.75	1.57	0.96	0.65	1.30	1.03	0.69	1.42	0.90	0.58	1.23	0.93	0.60	1.28	0.97	0.67	1.29	1.06	0.73	1.42
19	0.87	0.58	1.18	1.01	0.64	1.42	0.83	0.52	1.15	0.89	0.57	1.27	0.77	0.47	1.09	0.80	0.49	1.13	0.86	0.58	1.18	0.95	0.62	1.29
20	0.63	0.36	0.93	0.74	0.41	1.11	0.58	0.31	0.90	0.63	0.33	0.97	0.54	0.26	0.84	0.55	0.27	0.86	0.63	0.36	0.93	0.69	0.40	1.02
21	1.13	0.78	1.49	1.32	0.86	1.80	1.05	0.68	1.41	1.12	0.72	1.55	0.99	0.62	1.36	1.01	0.65	1.39	1.13	0.78	1.49	1.24	0.86	1.64
22	0.69	0.39	1.04	0.81	0.42	1.26	0.62	0.32	0.98	0.66	0.31	1.04	0.57	0.26	0.92	0.59	0.28	0.95	0.69	0.37	1.04	0.76	0.42	1.15
23	1.47	1.15	1.82	1.70	1.25	2.25	1.42	1.07	1.77	1.52	1.11	1.94	1.36	0.97	1.70	1.39	1.01	1.76	1.47	1.14	1.81	1.59	1.21	2.00
24	1.37	0.99	1.75	1.58	1.07	2.12	-	-	-	-	-	-	-	-	-	-	-	-	1.36	0.98	1.73	1.49	1.06	1.92
25	1.11	0.68	1.52	1.27	0.75	1.85	1.06	0.60	1.51	1.13	0.62	1.63	0.99	0.53	1.46	1.01	0.54	1.52	1.10	0.67	1.51	1.20	0.73	1.68
26	0.65	0.36	1.01	0.76	0.40	1.23	0.63	0.30	1.01	0.68	0.32	1.13	0.58	0.27	0.97	0.60	0.28	1.00	0.65	0.37	0.98	0.71	0.39	1.09
27	1.29	0.66	1.82	1.47	0.75	2.18	1.16	0.53	1.75	1.24	0.56	1.89	1.12	0.50	1.74	1.17	0.49	1.80	1.28	0.67	1.82	1.39	0.73	2.01
28	1.41	0.91	1.89	1.61	1.01	2.27	1.26	0.77	1.80	1.36	0.79	1.94	1.23	0.72	1.81	1.29	0.75	1.86	1.41	0.92	1.89	1.52	0.98	2.07
29	0.66	0.3	1.11	0.76	0.34	1.30	0.59	0.24	1.06	0.64	0.26	1.13	0.57	0.22	1.06	0.59	0.23	1.09	0.66	0.31	1.12	0.72	0.33	1.20
30	0.4	0.18	0.7	0.46	0.19	0.81	0.36	0.14	0.67	0.39	0.15	0.74	0.35	0.13	0.68	0.36	0.12	0.71	0.40	0.18	0.70	0.43	0.19	0.76
31	1.77	1.4	2.24	2.16	1.53	2.56	1.72	1.34	2.20	1.98	1.39	2.34	1.67	1.28	2.17	1.75	1.30	2.23	1.77	1.41	2.25	2.12	1.50	2.41
32	1.51	1.18	1.81	1.92	1.59	2.18	-	-	-	-	-	-	-	-	-	-	-	-	1.51	1.20	1.82	-	-	-
33	1.18	0.85	1.58	1.44	0.99	1.85	1.08	0.76	1.44	1.17	0.79	1.62	1.02	0.70	1.39	1.06	0.72	1.47	1.19	0.87	1.60	1.33	0.94	1.75
34	0.67	0.41	1	0.81	0.47	1.21	0.64	0.36	0.95	0.69	0.39	1.07	0.60	0.33	0.92	0.63	0.35	0.97	0.68	0.40	1.00	0.75	0.44	1.12
35	0.49	0.24	0.79	0.57	0.27	0.95	0.43	0.19	0.74	0.47	0.21	0.82	0.40	0.16	0.71	0.42	0.17	0.74	0.49	0.24	0.80	0.54	0.26	0.88
36	0.21	0.09	0.38	0.25	0.11	0.47	0.20	0.07	0.39	0.22	0.08	0.43	0.19	0.06	0.37	0.20	0.06	0.40	0.21	0.09	0.38	0.23	0.09	0.43
37	0.92	0.62	1.27	1.10	0.72	1.51	0.82	0.52	1.16	0.90	0.55	1.28	0.77	0.47	1.12	0.80	0.50	1.17	0.93	0.62	1.27	1.03	0.68	1.42
38	0.61	0.35	0.9	0.72	0.41	1.07	0.53	0.28	0.82	0.57	0.30	0.90	0.49	0.25	0.79	0.51	0.26	0.82	0.61	0.36	0.91	0.67	0.39	1.00
40	0.34	0.15	0.6	0.40	0.17	0.72	0.29	0.12	0.56	0.32	0.12	0.62	0.27	0.09	0.54	0.29	0.10	0.58	0.34	0.15	0.60	0.37	0.16	0.67
41	1.4	0.98	1.8	1.53	1.09	1.93	-	-	-	-	-	-	-	-	-	1.28	0.79	1.76	1.41	0.98	1.79	1.48	1.06	1.87
42	1.1	0.66	1.56	1.20	0.72	1.65	-	-	-	-	-	-	-	-	-	-	-	-	1.10	0.65	1.54	1.16	0.71	1.62
43	2	1.56	2.43	2.35	1.73	3.03	1.85	1.46	2.25	1.97	1.48	2.50	1.80	1.38	2.19	1.86	1.41	2.30	2.00	1.57	2.41	2.18	1.71	2.72
44	1.75	1.34	2.18	2.05	1.47	2.72	1.59	1.19	1.98	1.70	1.23	2.22	1.54	1.12	1.93	1.59	1.16	2.04	1.75	1.33	2.16	1.91	1.44	2.43
45	1.58	1.19	1.98	1.85	1.32	2.47	1.42	1.05	1.79	1.51	1.07	2.00	1.36	0.98	1.74	1.40	1.02	1.84	1.58	1.20	1.97	1.71	1.30	2.21
46	1.36	0.99	1.76	1.60	1.09	2.16	1.20	0.84	1.57	1.28	0.87	1.75	1.13	0.78	1.51	1.17	0.79	1.58	1.37	0.99	1.77	1.50	1.07	1.97
47	0.95	0.65	1.31	1.12	0.72	1.60	0.85	0.55	1.19	0.91	0.57	1.31	0.79	0.50	1.15	0.82	0.50	1.18	0.96	0.64	1.31	1.05	0.70	1.46
48	0.37	0.22	0.58	0.44	0.26	0.68	0.34	0.19	0.52	0.36	0.20	0.57	0.32	0.17	0.50	0.33	0.18	0.54	0.38	0.23	0.57	0.41	0.24	0.62
49	0.17	0.07	0.31	0.20	0.08	0.37	0.15	0.06	0.28	0.17	0.06	0.31	0.14	0.05	0.27	0.15	0.05	0.29	0.17	0.07	0.31	0.19	0.08	0.34
50	0.3	0.16	0.47	0.35	0.18	0.57	0.26	0.12	0.43	0.28	0.14	0.47	0.24	0.11	0.41	0.25	0.11	0.43	0.30	0.16	0.47	0.33	0.17	0.51
51	1.34	0.89	1.78	1.57	1.00	2.20	1.19	0.76	1.61	1.27	0.78	1.78	1.12	0.69	1.54	1.16	0.70	1.62	1.34	0.89	1.77	1.46	0.97	1.97
52	0.25	0.1	0.47	0.29	0.11	0.55	0.23	0.08	0.46	0.25	0.09	0.50	0.22	0.07	0.46	0.23	0.08	0.48	0.25	0.10	0.47	0.27	0.11	0.51
53	1.38	0.71	1.99	1.65	0.83	2.48	1.23	0.60	1.83	1.33	0.62	2.02	1.16	0.51	1.77	1.20	0.55	1.87	1.39	0.73	2.02	1.52	0.80	2.26
54	0.13	0.04	0.25	0.15	0.05	0.30	0.12	0.04	0.26	0.13	0.04	0.28	0.11	0.03	0.26	0.12	0.03	0.27	0.13	0.04	0.25	0.14	0.05	0.28
55	1.4	0.83	2.02	1.65	0.94	2.50	1.27	0.71	1.93	1.39	0.76	2.14	1.23	0.64	1.90	1.27	0.68	2.00	1.39	0.84	2.00	1.54	0.91	2.27
56	0.63	0.3	1.07	0.75	0.35	1.30	0.56	0.25	0.99	0.61	0.25	1.10	0.53	0.21	0.99	0.55	0.21	1.03	0.63	0.30	1.05	0.70	0.33	1.18
57	0.04	0.01	0.11	0.05	0.01	0.13	0.04	0.01	0.11	0.05	0.01	0.12	0.04	0.01	0.11	0.04	0.01	0.11	0.04	0.01	0.11	0.05	0.01	0.12

Additional File 4.5 Estimated Ages of nodes found in the Bayesian consensus trees for each analyses. Nd - node number; md - median node age; lo - lower boundary of the 95% highest-probability density; up - upper boundary of the 95% highest-probability density.

taxon	# species	# strains	accession number	taxon	# species	# strains	accession number
Acaryochloris sp. JJ8A6	5	14	AM710387	***Filamentous thermophilic cyanobacterium***	6	17	DQ471441
Synechococcus lividus C1*			AF132772	*Synechococcus sp.* C9*			AF132773
Thermosynechococcus elongatus BP-1*			BA000039	**G40R**	17	247	
Anabaena sp. PCC 7108	28	459	AJ133162	*Nodularia sp.* PCC 7804*			AJ133181
Arthronema gygaxiana UTCC 393	5	57	AF218370	***Geitlerinema sp.*** BBD HS217	8	97	EF110974
Pseudanabaena sp. PCC 7304*			6AF132933	***Gloeobacter violaceus*** PCC 7421	1	1	BA000045
Pseudanabaena sp. PCC 6802*			AB039016	***Gloeothece sp.*** PCC 6909/1	3	13	EU499305
Phormidium mucicola IAM M-221*			AB003165	***Halospirulina sp.***	2	6	NR_026510
Arthrospira platensis PCC 8005	9	42	X70769	***Leptolyngbya sp.*** ANT.LH52.1	6	332	AY493584
Lyngbya aestuarii PCC 7419*			AB075989	*Planktothrix sp.* FP1*			EU078515
Synechococcus sp. CC9605	13	567	AY172802	*Plectonema sp.* F3*			AF091110
Synechococcus sp. WH8101*			AF001480	***Microcoleus chthonoplastes*** PCC 7420	19	434	AM709630
Prochlorococcus sp. MIT9313*			AF053399	*Symploca sp.*PCC 8002*			AB039021
Cyanobium sp. JJ23-1*			AM710371	***Nostoc sp.*** PCC 7120	25	649	X59559
Synechococcus elongatus PCC 6301*			AP008231	***Oscillatoria sancta*** PCC 7515	30	129	AF132933
*Prochlorothrix hollandica**			AJ007907	*Trichodesmium erythraeum* IMS 101*			AF013030
Calothrix sp. PCC 7103	6	226	AM230700	***Oscillatoria sp.***	20	147	AJ133106
Chamaesiphon subglobosus PCC 7430	1	2	AY170472	***Prochloron sp.***	4	6	X63141
Chlorogloeopsis sp. PCC 7518	18	136	X68780	*Synechocystis sp.* PCC 6308*			AB039001
Fischerella sp. PCC 7414*			AB075986	***Radiocystis sp.*** JJ30-3	2	2	AM710389
Chroococcidiopsis sp. CC2	3	51	DQ914864	***Synechocystis sp.*** PCC 6803	4	51	NC_000911
Chroococcus sp. JJCM	10	325	AM710384	***Scytonema sp.*** U-3-3	4	45	AY069954
Microcystis aeruginosa strain 038*			DQ363254	***Spirulina sp.*** PCC 6313	6	29	X75045
Crinalium magnum SAG 34.87	2	2	AB115965	***Symphyonema sp.*** strain 1517	7	9	AJ544084
Starria zimbabweensis SAG 74.90*			AB115962	***Synechococcus sp.*** P1	1	2	AF132774
Cyanothece sp. PCC 8801	4	22	AF296873				
Dactylococcopsis sp.	4	34	AJ000711				
Dermocarpa sp. MBIC10768	8	41	AB058287				
Myxosarcina sp. PCC 7312*			AJ344561				
Myxosarcina sp. PCC 7325*			AJ344562				
Pleurocapsa sp. CALU 1126*			DQ293994				
Pleurocapsa sp. PCC 7516*			X78681				
*Dermocarpella incrassata**			AJ344559				

Additional File 4.6: Names of taxa and number of species and strains they can be assigned to in the cyanobacterial phylum. * Clades that have been pruned for the diversification rate analyses

Additional File 4.7 - Results of clade specific diversification rate estimation for eight different analyses.

	species				strains			
	clade	r	ϵ	aicc	clade	r	ϵ	aicc
1	root node 3 node 33	0.17 1.63 1.37	4.73E-8 9.07E-9 0.93	206.6	root node 3 node 34	0.28 2.39 1.08	7.79E-7 0.8 0.998	357.7
2	root node 3 node 33	0.38 1.43 0.99	1.14E-8 2.31E-9 0.95	218.4	root node 3 node 34	0.23 1.91 0.78	3.80E-7 0.83 0.999	371.2
3	root node 3 node 33	0.17 1.63 0.9	4.67E-7 2.29E-7 0.97	212.6	root node 3 node 34	0.28 1.99 1.06	1.07E-6 8.86E-001 0.999	365.4
4	root node 3 node 33	0.15 1.53 0.7	5.03E-7 1.61E-7 0.97	216.4	root node 3 node 34 split 4	0.25 1.9 0.94 0.75	1.79E-7 0.89 0.999 3.69E-5	366.8
5	root node 3 node 33	0.17 1.68 0.86	1.42E-7 3.22E-8 0.97	213.5	root node 3 node 34	0.29 .88 11 1.06	5.40E-7 0.9 0.999	368.1
6	root node 3 node 33	0.16 1.67 0.81	1.24E-6 4.09E-7 0.97	212.4	root node 3 node 34	0.27 1.97 1.01	3.49E-7 0.9 0.999	367.9
7	root node 3 node 33	0.17 1.63 1.38	3.90E-10 2.45E-7 0.93	206.7	root node 3 node 34	0.28 2.39 1.07	3.90E-7 0.8 0.998	357.8
8	root node 3 node 33	0.15 1.47 0.74	4.42E-7 1.83E-7 0.97	216.3	root node 3 node 34	0.24 1.94 0.89	7.89E-7 0.85 0.999	370.0

Additional File 4.7: Results of clade specific diversification rate estimation for eight different analyses.

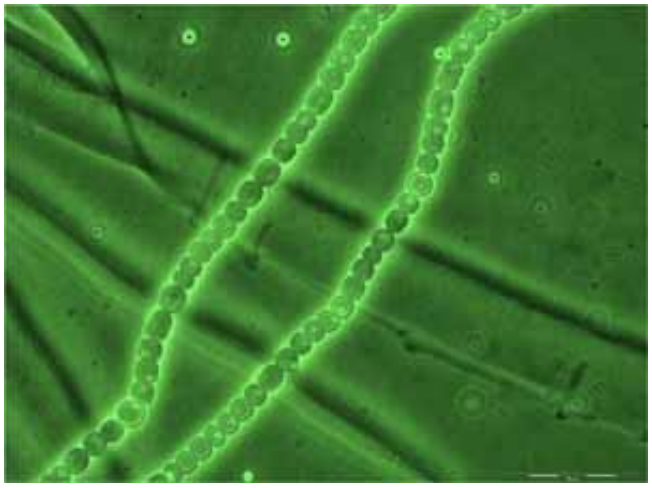
Additional File 4.8 - Nucleotide sequence of G40R

A cyanobacterial strain isolated from North Sea in 2008.

>G40R

GGGAACTGTGGCTAATACCGGATATGCCGAGAGGTGAAAGCTAGGCCTG
AAGATGAGCTCGCGTCTGATTAGCTAGTAGGTGTGGTAAGAGCGCACCT
AGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGAC
TGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTCCG
CAATGGGCGAAAGCTGACGGAGCAATACCGCGTGAGGGAGGAAGGCTCT
TGGGTTGTAAACCTCTTTTCTCAAGGAAGTGACGGTACTTGAGATAAGC
ATCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGATGCAAGCG
TTATCCGGAATGATTGGGCGTAAAGGGTCCGCAGTGGCTGTGTAAGTCT
GCTGTTAAAGAATCTAGCTTAACTAGATAAAGCAGTGGAAACTACATAG
CTAGAGTGCGTTCGGGGTAGAGGGAATTCCTGGTGTAGCGGTGAAATGC
GTAGATATCAGGAAGAACACCAGTGGCGAAGCCTCTACTAGGCCGCACT
GACACTGAGGACGAAAGCTAGGGAGCGAATGGATTAGATACCCCAGTAG
TCCTAGCCGTAAACGATGGATACTAGGCGTGGGACCCGAGCCGTGCCGG
AGCTAACGCGTTAAGTATCCCGCCTGGGGAGTACGCACGCAAGTGTGAA
ACTCAAAGGAATTGACGGGGGCCGCACAAGCGGTGGAGTATGTGGTTAA
TTCGATGCAACGCGAAGAACCTTACCAAGCTTGACATGTCGCGAATCTT
CTTGAAAGGGAAGAGTGCTTCGGAGCGCGAACACAGGTGGTGCATGGCT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGC
AACCCTCGTTTTTAGTTGCAGCATTAAGTTGGGCACTCTAGAGAGACTG
CCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCAGCATGCCC
CGTCTTGGGCTACACACGTACTACAATGCTACGGACAAAGGGCAGCTAC
ACAGCAATGTGATGCAAATCTGAAACCGTAGCTCAGTTCAGATCGCAG

Additional publications



The evolutionary path to terminal differentiation and division of labor in cyanobacteria

Authors:
Valentina ROSSETTI, Bettina E. SCHIRRMEISTER , Marco V. BERNASCONI and Homayoun C. BAGHERI

Journal of Theoretical Biology 262 (2010) 23–34

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

The evolutionary path to terminal differentiation and division of labor in cyanobacteria

Valentina Rossetti [a], Bettina E. Schirrmeister [a], Marco V. Bernasconi [b], Homayoun C. Bagheri [a,*]

[a] Institute of Zoology, University of Zurich, Zurich, Switzerland
[b] Zoological Museum, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Article history:
Received 15 December 2008
Received in revised form
21 August 2009
Accepted 6 September 2009
Available online 15 September 2009

Keywords:
Multicellularity
Tragedy of the commons
Cooperation
Phytoplankton
Circadian rhythm

ABSTRACT

A common trait often associated with multicellularity is cellular differentiation, which is a spatial separation of tasks through the division of labor. In principle, the division of labor does not necessarily have to be constrained to a multicellular setting. In this study, we focus on the possible evolutionary paths leading to terminal differentiation in cyanobacteria. We develop mathematical models for two developmental strategies. First, of populations of terminally differentiated single cells surviving by the exchange of common goods. Second, of populations exhibiting terminal differentiation in a multicellular setting. After testing the two strategies against the effect of disruptive mutations (i.e. "cheater" mutants), we assess the effects of selection on the optimization of the ratio of vegetative (carbon fixing) to heterocystous (nitrogen fixing) cells, which in turn leads to the maximization of the carrying capacity for the population density. In addition, we compare the performance of differentiated populations to undifferentiated ones that temporally separate tasks in accordance to a day/night cycle. We then compare some predictions of our model with phylogenetic relationships derived from analyzing 16S rRNA sequences of different cyanobacterial strains. In line with studies indicating that group or spatial structure are ways to evolve cooperation and protect against the spread of cheaters, our work shows that compartmentalization afforded by multicellularity is required to maintain the vegetative/heterocyst division in cyanobacteria. We find that multicellularity allows for selection to optimize the carrying capacity. These results and the phylogenetic analysis indicates that terminally differentiated cyanobacteria evolved after undifferentiated species. In addition, we show that, in regimes of short daylight periods, terminally differentiated species perform worse than undifferentiated species that follow the day/night cycle; indicating that undifferentiated species have an evolutionary advantage in regimes of short daylight periods.

1. Introduction

1.1. Multicellularity and the germline–soma divide

Multicellular organisms undergo cellular differentiation in order to perform distinct tasks. A fundamental example is differentiation into germline and somatic cells. This division of labor was first elucidated by Weismann (1889) upon studying aquatic animals such as hydrozoans, and green algae of the order Volvocales (Schleip, 1934). He distinguished between germ cells (*Keimzellen*) that contribute cells and hereditary material to the subsequent generation of a multicellular individual, and somatic cells (*Somatische Zellen*) that help in the survival of an individual during its lifetime. In some animals, differentiation into germ cells can be irreversible, referred to as "terminal differentiation." The germline–soma divide is now viewed as a fundamental organizational scheme in complex multicellular organisms, and is central to understanding the interplay between natural selection at the level of the multicellular individual, and competition between its component cells (Buss, 1983, 1988).

The separation between a germline and soma is not unique to Eukaryotes, and is also mirrored in differentiated multicellular cyanobacteria (Saier and Jacobson, 1984). The latter can differentiate into vegetative and heterocystous cells, which are functionally equivalent to germline and soma, respectively. Moreover, differentiation into heterocystous cells is terminal. The fact that the same fundamental organizational scheme for the division of labor has independently appeared in such disparate lineages suggests that there may be general conditions that favor the emergence of such an organization. With this view in mind, multicellular cyanobacteria can serve as a model organism for understanding the developmental and ecological conditions that lead to the evolution of terminal differentiation and a germline–soma divide.

* Corresponding author.
E-mail addresses: valentina.rossetti@zool.uzh.ch (V. Rossetti), bagheri@zool.uzh.ch (H.C. Bagheri).

0022-5193/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2009.09.009

Although there is a growing literature on modelling the ecology and population dynamics of nitrogen fixing cyanobacteria (Tilman, 1977; Roussel et al., 2000; Rabouille et al., 2006; Agawin et al., 2007), the factors that can affect the evolution of multicellularity and differentiation in these organisms has not been examined. In this work we try to approach several fundamental questions. First, we ask what are the fundamental conditions necessary for the evolutionary stability of a terminally differentiated soma in cyanobacteria. Second, we ask how differentiation is related to fitness, and how the rate of differentiation can be optimized in an evolutionary context. Third, we address some of the ecological conditions that may favor the spatial vs. temporal separation of tasks between cyanobacterial cells. Fourth, we examine the phylogenetic history of cyanobacteria in light of our theoretical results.

In the rest of this introduction we discuss the empirical and theoretical background necessary for the models that we subsequently develop.

1.2. Multicellularity in cyanobacteria

The cyanobacteria encompass both unicellular and multicellular species, and are among the most ancient multicellular organisms known (Schopf, 1994). Among multicellular species, differentiation into heterocystous forms seems to have a monophyletic origin (Turner et al., 1999; Seo and Yokota, 2003; Tomitani et al., 2006). Multicellular cyanobacteria such as members of the genera *Anabaena* and *Nostoc* are often present as filaments differentiated into two kinds of cells: vegetatives and heterocysts (Wolk, 1996). Some species also have akinete cells specialized for surviving harsh conditions (hence being similar to spores in their function). We will not deal with akinetes in this study. Vegetative cells are photosynthetic and reproduce by cell division, giving rise to either vegetative or heterocystous cells. They use solar energy and carbon dioxide for the purpose of carbon fixation, and fixed nitrogen in the form of nitrates for building molecules such as amino acids. Fixed nitrogen is produced by heterocysts, whose main task is nitrogen fixation using free atmospheric nitrogen. Heterocysts cannot divide and originate from the division of vegetative cells (a portion of vegetative divisions leads to heterocysts instead of vegetative cells). The need for division of labor between cells that either fix nitrogen or carbon arises from inhibitory chemical interactions between photosynthesis and nitrogen fixation. By having the two chemical reactions occur in different cells, filamentous cyanobacteria can improve the efficiency of nitrogen fixation. In undifferentiated cyanobacteria such as *Synechocystis* sp. or *Oscillatoria* sp., the main strategy is to have a day and night cycle (circadian rythm) (Stal and Krumbein, 1987; Kondo et al., 1993; Bergman et al., 1997; Kageyama et al., 2006; Kurosawa et al., 2006), according to which photosynthesis and nitrogen fixation are temporally separated. The interactions among vegetatives and heterocysts can be also framed in the context of cooperation. Heterocysts sacrifice the possibility of reproduction and fix nitrogen for all the cells, in this sense being a fully altruistic entity. Vegetative cells are also cooperative: they do not use all their progeny to pass their genes to the next generation, because part of it will become heterocystous and will lose this ability. If vegetative cells produce few or no heterocysts in order to maximize their reproductive success, they act as defectors.

A detailed classification of the cyanobacteria has been made by Rippka et al. (1979). Cyanobacteria are phenotypically classified into five sections (I–V), which are schematically depicted in Fig. 1. In the case of heterocystous section IV species such as *Anabaena* sp., it has been recently established that filaments are truly

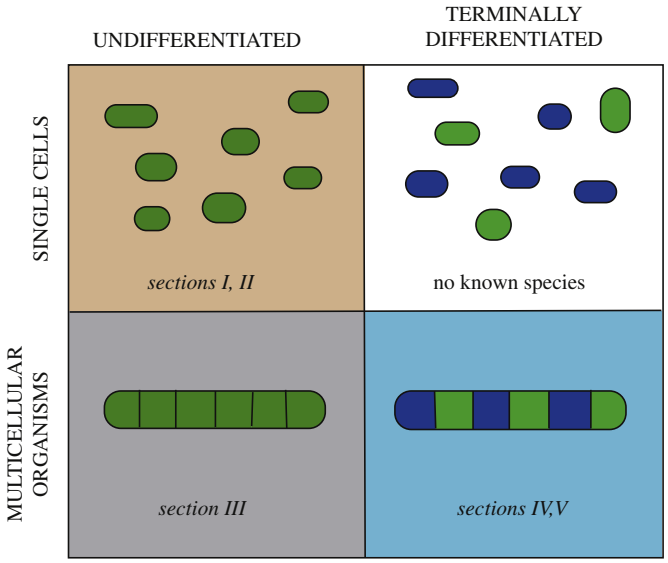


Fig. 1. Schematic classification of cyanobacterial species based on Rippka et al. (1979).

multicellular, in the sense that the periplasmic space along the filament is continuous (Flores et al., 2006; Mariscal et al., 2007). This allows for an exchange conduit for nutrients and other molecules between cells. Given that cyanobacteria are gram negative and possess two membranes, the continuity of the periplasm is achieved via the outer membrane, which forms a unified compartment around a chain of cells, rather than individual cells. Each cell is in turn also encapsulated by its own cytoplasmic membrane. In addition, there is some evidence for direct exchange between the cytoplasms of adjacent cells through membrane channels (Mullineaux et al., 2008).

1.3. Evolution of multicellularity and cooperation

The evolutionary transition between unicellular and multicellular forms involves conflicts between different levels of selection (Buss, 1988; Smith and Szathmáry, 1995; Rainey and Rainey, 2003; Michod, 2007). The benefits associated with multicellularity may for example include size, nutritional advantages, collective protection against antagonists, and division of labor (Shapiro, 1998; Bonner, 2000; Kaiser, 2001). However, multicellular organization does not automatically imply the existence of differentiation. Undifferentiated multicellularity can have its own advantages over single-celled organization (Pfeiffer et al., 2001; Pfeiffer and Bonhoeffer, 2003; Willensdorfer, 2009).

Once multicellularity has evolved, one can consider the conditions under which cellular differentiation would be advantageous. For example, the division between germline and soma can be analyzed as a consequence of the interplay between two fitness components, namely reproduction and survival (Weismann, 1889; Michod et al., 2006). Cooperation among cells is fundamental in building a differentiated multicellular organism. Single entities lose the opportunity of selfish reproduction in order to become part of a community of cells. They produce and share nutrients with the others instead of using everything to their advantage, hence increasing the fitness of the multicellular unit (Michod and Roze, 2001). However, such a behavior can be abandoned by defectors (or cheaters), who exploit the cooperative acts but do not contribute to the common good. Following the work of Hamilton (1964a, 1964b), various studies have been made about cooperation and selfish behavior using game theoretic approaches (Smith and Price, 1973; Hofbauer et al., 1979; Smith,

1982; Hofbauer and Sigmund, 1998; Ohtsuki et al., 2006). Non-cooperative or "cheating" behavior is common in many ecosystems: cheaters can exhibit selective advantages over the competitors (Axelrod and Hamilton, 1981; Sachs et al., 2004; Boomsma and Franks, 2006), but can lead to reciprocal extinction or to stable mutualistic associations (Doebeli and Knowlton, 1998; Roberts and Sherrat, 1998; Ferriere et al., 2002). Over-exploitation of a common good by cheaters is often referred to as the "tragedy of the commons" (Hardin, 1968). It is known that some kind of subpopulation grouping is required for resolving this problem. The classic explanations are kin selection (Hamilton, 1964a, 1964b; Smith, 1964; Frank, 1994; Lehmann and Keller, 2006; West et al., 2006) and reciprocity (Trivers, 1971; Axelrod and Hamilton, 1981; Leimar and Hammerstein, 2001; Hammerstein, 2003; Lehmann and Keller, 2006; Suzuki and Akiyama, 2008). Other mechanisms are for example differential dispersal(Enquist and Leimar, 1993; Hochberg et al., 2008), resource supply (Brockhurst et al., 2008), spatial structuring of the population (Nowak and May, 1992; Nowak et al., 1994; Ferriere and Michod, 1996; Nakamaru et al., 1997; Pfeiffer et al., 2001; Pfeiffer and Bonhoeffer, 2003), allowing for the random emergence of association groups (Michod, 1983; Szathmáry and Demeter, 1987; Killingback et al., 2006), or imposing threshold conditions in the rules of the game (Bach et al., 2006). Various aspects of these theories have been validated in microbes (Buss, 1982; Strassmann et al., 2000; Velicer et al., 2000; Rainey and Rainey, 2003; Griffin et al., 2004; Travisano and Velicer, 2004; West et al., 2006). For example, assortment and phenotypic noise can allow the evolution of self-destructive-cooperation in *Salmonella thyphimurium* (Ackermann et al., 2008), while kin selection limits cheating in the slime mold *Dictyostelium* spp. (Buss, 1982; Gilbert et al., 2007).

Hypercycles, which are autocatalytic networks of enzyme reactions are another system where the issue of cheating and the importance of population subdivision arises (Eigen, 1971; Eigen and Schuster, 1977, 1978). Hypercycles are susceptible to invasion by "parasitic" enzymes that have reduced catalytic activity for the replication of their target enzyme. It has been suggested several times (Eigen, 1971; Eigen and Schuster, 1978; Smith, 1979; Eigen et al., 1980; Michod, 1983) that one way to escape the problem of parasite invasion in the latter case would be the evolution of compartments or "protocells" that allow different hypercycles to compete. The "stochastic corrector model" of Szathmáry and Demeter (1987) implements a version of this concept (Smith and Szathmáry, 1995). In a similar vein, an alternative path to achieve population substructuring is the introduction of spatial heterogeneity (Boerlijst and Hogeweg, 1991; Attolini and Stadler, 2006; Sardanyés and Solé, 2006; Fontanari et al., 2006; Hogeweg and Takeuchi, 2003).

2. Methods

At present, there are no known single-celled species of cyanobacteria that terminally differentiate to form collaborative single species consortia as a means to divide labor between nitrogen and carbon fixers (top-right box in Fig. 1). We model the latter hypothetical scenario (single-celled model) and that of differentiated multicellularity (compartmental model, bottom-right box in Fig. 1).

2.1. Mathematical models

2.1.1. The single-celled model

We consider a single-celled model (Fig. 2a) where vegetatives, heterocysts and cheater vegetatives compete for nitrate, fixed carbon and solar energy. The vegetative cells convert the solar energy into chemical energy (fixed carbon), while the nitrate is produced by heterocysts. Vegetative cells divide into vegetative and heterocyst cells in different proportions. The cheaters, when present in the system, produce and consume resources at the same rate as the non-cheater vegetatives, but they produce less heterocysts—or do not produce them at all. In this model, the resources are shared by all cells living in the environment.

We describe the single-celled model with the following system of ODEs:

$$\frac{dN}{dt} = 2aH\frac{C}{C+k} - p_3N - r(V+V')Z - q\frac{N}{N+k}(V+H+V')$$

$$\frac{dV}{dt} = -p_3V + p_vVZ$$

$$\frac{dH}{dt} = -p_3H + p_hVZ + p_{h'}V'Z$$

$$\frac{dV'}{dt} = -p_3V' + p_{v'}V'Z$$

$$\frac{dC}{dt} = c_eZ_L - p_3C - r(V+V')Z - q\frac{C}{C+k}(V+H+V') - aH\frac{C}{C+k} \qquad (1)$$

where

$$Z = Z(N,C) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_CC} + \frac{1}{k_NN} + \frac{1}{k_{NC}NC}}$$

$$Z_L = Z_L(I,G) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_II} + \frac{1}{k_GG} + \frac{1}{k_{IG}IG}}$$

I = irradiance (constant)

$G = V + V'$ (photosynthetic units)

$$p_v + p_h = 1, \quad p_{v'} + p_{h'} = 1 \qquad (2)$$

In (1), N is the nitrate concentration (mol/cm^3), V, H and V' are the concentrations of resident vegetative, heterocyst and cheaters cells, respectively (cells/cm^3) and C (mol/cm^3) the concentration of chemical energy (sugar in the form of glucose).

The equations have been built on and can be explained by the following assumptions:

(i) *Reproduction and housekeeping*: The parameter p_v indicates the proportion of vegetative cells originated at any reproduction event. Cheater cells behave similar to vegetative cells, except for their p_v value, designated as $p_{v'}$. Heterocysts are produced in proportion p_h by vegetatives and in proportion $p_{h'}$ by cheaters. Vegetative cells use chemical energy to support dividing ($-rVZ$) and living costs ($-qV(C/C+k)$). Heterocyst cells use chemical energy for nitrogen fixation ($-aH(C/C+k)$) and living costs ($-qH(C/C+k)$). We assume the same death rate p_3 for all cells. The reproduction of vegetatives is regulated by $Z = Z(N,C)$, a Michaelis–Menten type saturation function for two substrates (here nitrogen and sugar).

(ii) *Energetics of nitrogen fixation*: Heterocysts are responsible for nitrogen fixation, which requires about 19 molecules of ATP. Taking into account that a molecule of glucose gives roughly 38 ATP molecules (Lawlor, 1990), we have

$$\frac{38\text{ATP/glucose}}{19\text{ATP/fixed N}} \simeq 2 \cdot \text{fixed N/glucose}.$$

This relation is the basis for an assumed ratio of 2 molecules of nitrogen produced for every glucose consumed. Nitrogen

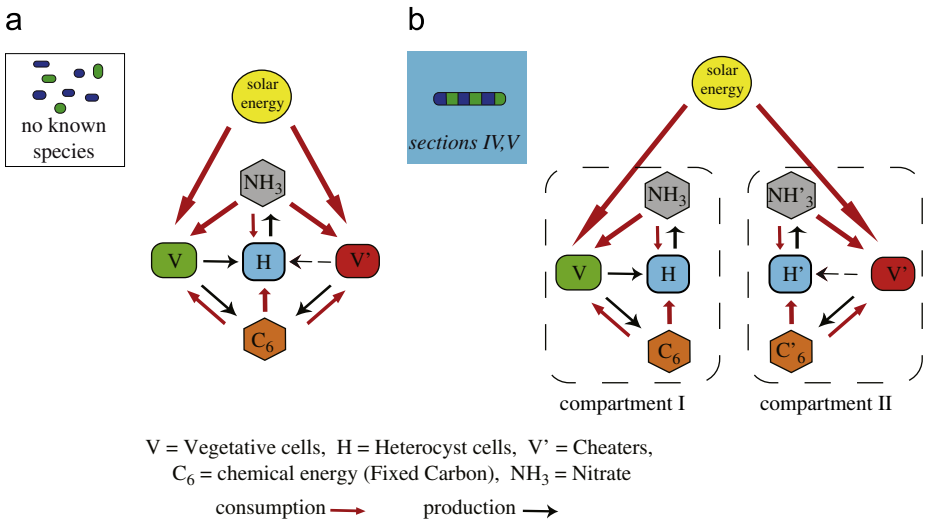


Fig. 2. Schematic representation of interactions between the variables of the models.

fixation is only possible whenever carbon (sugar) is available and it is limited by a saturation function dependent on sugar. It is assumed that free nitrogen is not a limiting factor. Decrease in nitrate is due to natural decay ($-p_3N$), to housekeeping or living cost ($-q(N/N+k)(V+H)$) and to reproduction of vegetative cells ($-rVZ$).

(iii) *Light harvesting*: The solar irradiance, I, is treated as a static parameter, as is common practice in basic models of photosynthesis (Han, 2001; Rubio et al., 2003). The irradiance is absorbed and transformed into chemical energy (six-carbon-sugar) by vegetative and cheater cells. The total production of sugar depends on I and on the total concentration of photosynthetic units G (PSU). We use the function $Z_L = Z_L(I, G)$ to describe the connection between light harvesting and sugar production. In Z_L, the solar irradiance I is absorbed by the photosynthetic units $G = (V+V')$ present in both the normal and cheater vegetative cells, hence I and G are considered as substrates for Z_L. This type of saturation function ensures that whenever a substrate is limiting the considered cellular activity, the potential increase of the other substrate does not enhance the activity. The solar energy is converted into carbon at a rate c_e. We assume that CO_2 concentrations are not a limiting factor. Carbon is subject to a natural decay ($-p_3C$).

(iv) *Carbon to nitrogen consumption ratios*: During exponential growth, the average ratio of carbon to nitrate is roughly C : N $\simeq$ 6 : 1 (Vrede et al., 2002) in a bacterial cell. The uptake of carbon should hence be 6 times higher than that of nitrate. However, considering that a molecule of glucose contains six carbon atoms, we have C : N $\simeq$ 1 : 1. The parameters r and q represent the rate of uptake of N and C for reproduction and housekeeping, respectively. Based on a 1:1 expectation for C:N content, the coefficients for C:N consumption for both reproduction and housekeeping are assumed to be one.

2.1.2. The compartmental model of multicellularity

In the compartmental model (Fig. 2b), only the solar energy is shared, while different compartments produce and consume their own sugar and nitrate units. The cheater is now an aggregate in which the proportion of vegetative cells produced at each division is higher than in the other. The compartmental model is a simplified representation of multicellularity, in which each compartment represents a distinct multicellular individual. Both compartments compete for the same solar energy source, but each has its own cells, nitrate and chemical energy. The following ODE set describes the dynamics of the compartmental model:

$$\frac{dN}{dt} = 2aH\frac{C}{C+k} - p_3N - rVZ - q\frac{N}{N+k}(V+H)$$

$$\frac{dV}{dt} = -p_3V + p_vVZ$$

$$\frac{dH}{dt} = -p_3H + p_hVZ$$

$$\frac{dC}{dt} = \frac{V}{V+V'}c_eZ_L - p_3C - rVZ - aH\frac{C}{C+k} - q\frac{C}{C+k}(V+H)$$

$$\frac{dN'}{dt} = 2aH'\frac{C'}{C'+k} - p_3N' - rV'Z' - q\frac{N'}{N'+k}(V'+H')$$

$$\frac{dV'}{dt} = -p_3V' + p_{v'}V'Z'$$

$$\frac{dH'}{dt} = -p_3H' + p_{h'}V'Z'$$

$$\frac{dC'}{dt} = \frac{V'}{V+V'}c_eZ_L - p_3C' - rV'Z' - aH'\frac{C'}{C'+k} - q\frac{C'}{C'+k}(V'+H') \quad (3)$$

where

$$Z = Z(N,C) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_CC} + \frac{1}{k_NN} + \frac{1}{k_{NC}NC}}$$

$$Z' = Z'(N,C) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_CC'} + \frac{1}{k_NN'} + \frac{1}{k_{NC}N'C'}}$$

$$Z_L = Z_L(I,G) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_II} + \frac{1}{k_GG} + \frac{1}{k_{IG}IG}} \quad (4)$$

I = irradiance (constant)

$G = V+V'$ (photosynthetic units)

$p_v + p_h = 1, \quad p_{v'} + p_{h'} = 1$

Variables N, V, H, C, respectively, represent nitrate, vegetatives, heterocysts and chemical energy concentrations of the first

filament, respectively, while N', V', H', C' represent the corresponding variables for the second compartment. The functions Z and Z' have the same meaning as in the single-celled model, except that in this case, they are functions of the respective nitrate and chemical energy of the two compartments. Competition for light between compartments is expressed in the function Z_L, by the variable $G = V + V'$. Light harvesting is due to both compartments, but the terms $V/V+V'$ and $V'/V+V'$ indicate that the income of sugar into the different aggregates is mediated by the concentration of photosynthetic units belonging to the corresponding compartments, hence allowing competition for light. The same assumptions (i)–(v) for system (1) listed in Section 2.1.1 hold also for system (3).

2.2. Numerical analysis of the models

Due to the high nonlinearity of the equations, we do not derive the analytical expression of the steady states of the system nor do we analytically carry out stability analysis. However, numerical simulations indicate that both systems can evolve towards three different kinds of steady state: one corresponding to the extinction of resident and cheater populations (Y_1), one in which only the resident population survives (Y_2), and one in which the cheaters overcome the resident population (Y_3). Parameters listed in Table 1 have been used in the simulations as default parameters for a low irradiance case. Structural stability of the models has been tested by random sampling of other parameter values and initial conditions in $\mathbb{R}^{21}$ and in $\mathbb{R}^{22}$ for the single-celled and compartmental model, respectively (results in Supplementary Information). Numerical integration has been performed using a variable order solver based on linear implicit multistep methods, implemented in function `ode15s` of Matlab (http://www.mathworks.com/).

2.3. Evolutionary stability against cheaters

Using the models in Section 2.1, we simulate competitions between a resident population and either pure or partial cheaters. Mutation is introduced in the systems in the following ways. In the case of pure cheaters, the latter are considered as the mutant. In the single-celled model, mutants are present in the mixed population from the beginning in a given proportion. In the compartmental model, they are introduced in only one of the compartments, while the other one is preserved. In the case of partial cheaters, in both models and for each mutational event, the strain with a p_v different from the resident strain is considered as the mutant.

2.3.1. Evolutionary optimization of vegetative/heterocyst ratio

We consider competitions between strains that differ in their p_v value, with $0 < p_v < 1$. We simulate consecutive competitions between a resident strain (wild type) and a newly arrived mutant. Each step of the simulation is a mutational event, in which after the competition, the winner strain establishes its p_v value as the wild type for the next generation (see Supporting Information for details of the algorithm).

2.4. Division of labor in time and space: periodic vs. differentiated cyanobacteria

We model a population of undifferentiated cyanobacteria subject to day/night irradiance cycle by the following ODE system:

$$\frac{dN}{dt} = 2a\delta_n \frac{C}{C+k}V - p_3 N - rVZ - q\frac{N}{N+k}V$$

$$\frac{dV}{dt} = -p_3 V + VZ$$

$$\frac{dC}{dt} = c_e Z_L - p_3 C - rVZ - (a\delta_n + q)\frac{C}{C+k}V \tag{5}$$

where

$$Z = Z(N,C) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_C C} + \frac{1}{k_N N} + \frac{1}{k_{NC} NC}}$$

$$Z_L = Z_L(I,V) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_I I} + \frac{1}{k_V V} + \frac{1}{k_{IV} IV}} \tag{6}$$

$$I(t) = A\frac{(\rho(t)+1)^\gamma}{(m^\gamma + (\rho(t)+1)^\gamma)} \tag{7}$$

$$\rho(t) = \sin\left(\frac{\pi t}{12}\right) \tag{8}$$

$$\delta_n = 1 - \frac{(\rho(t)+1)^\gamma}{(m^\gamma + (\rho(t)+1)^\gamma)} \tag{9}$$

In (5), N, C, and V are fixed nitrogen, carbon and cell concentrations, respectively. The irradiance I and the nitrogen

Table 1
Parameters values used in the displayed simulations.

Parameter description	Symbol	Value	Unit
Uptake of N and C for reproduction	r	5	mol cells^{-1}
Uptake of N and C for housekeeping	q	0.8	mol cells^{-1} s^{-1}
Uptake of C for N-fixation	a	1	mol cells^{-1} s^{-1}
Decay rate	p_3	0.001	s^{-1}
Irradiance	I	1000	μE cm^{-2} s^{-1}
Rate of energy conversion	c_e	0.8	mol cm^{-3}
Total stoichiometric concentration	r_0	0.1	
First order rate constant	k_0	10	s^{-1}
Nitrate specificity constant	k_N	10	mol^{-1} cm^3 s^{-1}
Carbon specificity constant	k_C	10	mol^{-1} cm^3 s^{-1}
N–C product specificity constant	k_{NC}	1	(mol^{-1} cm^3)^2 s^{-1}
Irradiance specificity constant	k_I	10	μE^{-1} cm^2
PSU specificity constant	k_G, k_V	10	cells^{-1} cm^3 s^{-1}
Irradiance-PSUproduct specificity constant	k_{IG}, k_{IV}	1	cells^{-1} μE^{-1} cm^5
Transformed Hill sine/cosine functions	γ	30	

Abbreviations: E, Einstein; PSU, photosynthetic units.

fixation function δ_n are Hill transformed sine curves that represent the daylight dependent periodicity (Marler et al., 2006). The default parameter values are as in Table 1. In this model, we assume that cells do not have a true internal circadian rhythm, but follow the external day/night alternation. During daylight, the periodic organisms only perform photosynthesis ($I \simeq 1, \delta_n \simeq 0$) because of O_2 inhibition. At night, when the absence of light impede photosynthesis, nitrogen fixation is allowed ($I \simeq 0, \delta_n \simeq 1$). We compare the performance of undifferentiated periodic species in (5) with the multicellular differentiated species, when the irradiance is described with a Hill transformed sine curve as in (7). To model a population of differentiated cyanobacteria subject to a day/night irradiance cycle, we modify system (1) by removing cheaters (V') and using (7) for irradiance:

$$\frac{dN}{dt} = 2a\frac{C}{C+k}H - p_3N - rVZ - q\frac{N}{N+k}(V+H)$$

$$\frac{dV}{dt} = -p_3V + p_vVZ$$

$$\frac{dH}{dt} = -p_3H + p_hVZ$$

$$\frac{dC}{dt} = c_eZ_L - p_3C - rVZ - aH\frac{C}{C+k} - q\frac{C}{C+k}(V+H) \qquad (10)$$

where Z, Z_L and I are as in (6) and (7). As nitrogen fixation is always performed by heterocysts, the function δ_n is not needed. We compare the performance of the two models with different day and night durations, by changing the value of m in I and δ_n. Large and small values of m correspond to long and short dark periods, respectively. We map the values of m into a percentage of daylight (details in Supplementary Information). As the mapping is based on an approximation of the duration of the day, it is not suited to treat neither complete darkness nor absence of darkness. For these cases, we directly set $I = 0$, 1, respectively.

2.5. Phylogenetic analysis of cyanobacteria

For this study, 16S rRNA gene sequences of 37 cyanobacteria and an outgroup were obtained from GenBank (Table 1 in Supporting Information). The ingroup is represented by nine single celled bacteria from clade I, four single celled bacteria from clade II, 14 multicellular bacteria from clade III, seven multicellular heterocyst forming bacteria from clade IV and four branching bacteria from clade V. Our labeling into clades I–V is based on Rippka et al. (1979). *Agrobacterium tumefaciens* was used for outgroup comparison as suggested in previous studies (Honda et al., 1999; Partensky et al., 1999). Details of the analysis are provided in Supporting Information.

3. Results

3.1. Effect of pure cheaters on evolutionary stability

In the single-celled model, introduction of pure cheaters leads to the extinction of the population (Fig. 3a). Cheaters grow faster and subtract resources from the resident population, which eventually starts decaying after reaching an initial peak. Once the normal vegetative cells are extinct, no entity in the system is able to produce nitrate and the cheaters also die. The collapse of the system in the single-celled model is primarily caused by the fact that the resources are shared between organisms.

In the compartmental model, a pure cheater cell gives rise to an aggregate of cells that cannot sustain itself. The compartmentalization afforded by separate multicellular aggregates (i.e. "multicellular" individuals) allows genetically related cells to protect their resources from a cheater invasion in another aggregate. Hence a cheater can destroy the multicellular aggregate that it arises in, but it cannot destroy the whole population (Fig. 3b). The basic dynamics of the models without cheaters are provided in Supplementary Information.

3.2. Effect of partial cheaters on evolutionary stability

Partial cheating refers to the situation in which a mutant vegetative cell produces heterocysts in a smaller proportion than the resident vegetative cells. We investigated the criterion that leads to the success of one genotype over the other, when two competing populations differ in their p_v value. We tested the outcomes of competitions in both models using Monte Carlo simulations, where pairs of p_v and $p_{v'}$ values were sampled randomly in the interval $[0, 1]$. Vegetative cells belonging to strains with p_v and $p_{v'}$ are indicated by V and V', respectively. Vegetative steady states after competitions are shown in Fig. 4. Fig. 4a shows the results for the single-celled model. In this case we find that at a steady state, $V > V'$ when $p_v > p_{v'}$ and similarly, $V' > V$ when $p_{v'} > p_v$. Hence we conclude that in the single-celled model, the winning factor in the competitions is the value of p_v. The strain with the higher p_v outcompete the other. This result holds when randomly sampling through alternative parameter values and initial conditions (see Supporting Information).

When tested for the compartmental model, the latter winning criterion does not hold. Fig. 4b shows that having a higher p_v does

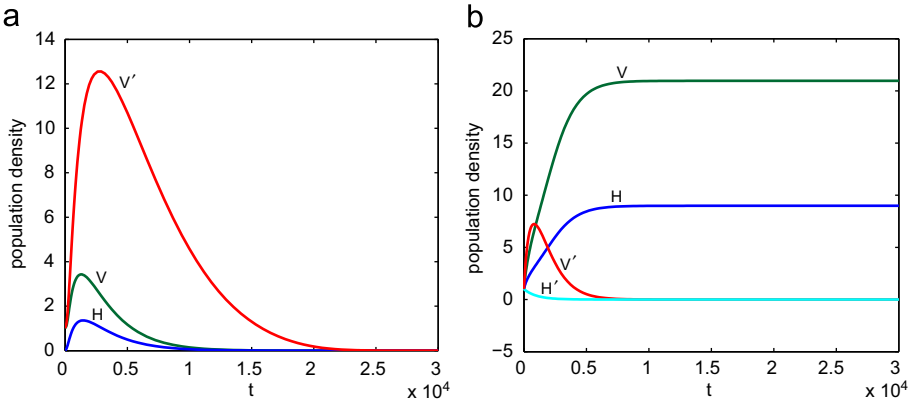


Fig. 3. Effect of pure cheaters on the resident population. V, vegetatives, H, heterocysts, V', cheaters, H', heterocysts in the compartment with cheaters. In (a), all the cell types reach zero, while in (b), cells of the non-mutant compartment can grow and reach a positive steady state.

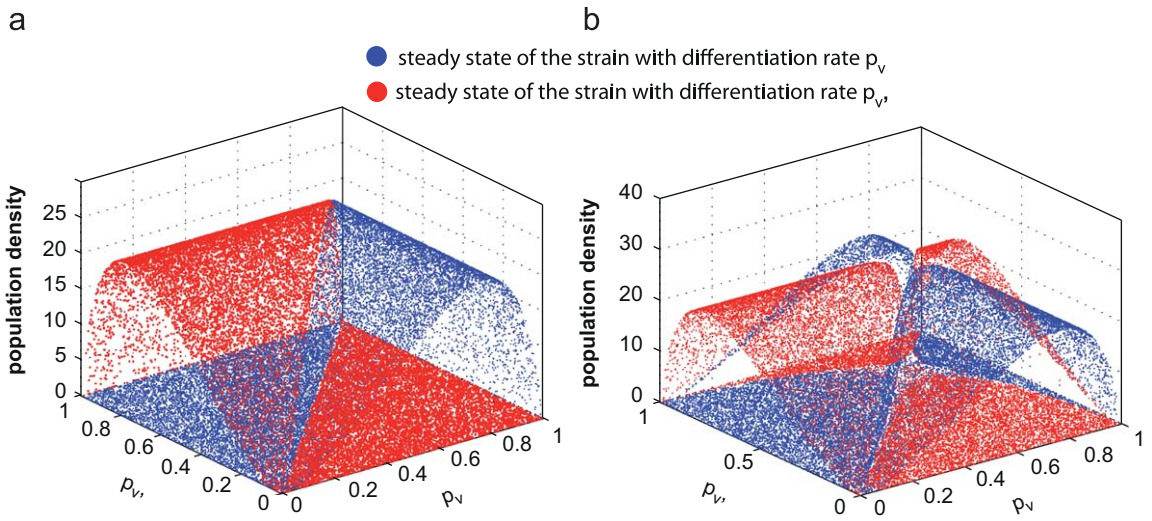


Fig. 4. Steady state density of vegetative cells after competitions between strains differing in their p_v value (plots show representative 15,000 competitions from 115,000). The blue and red dots correspond to strains competing with differentiation rates of p_v (blue) and $p_{v'}$ (red), respectively. (a) In the single-celled model, the strain with the higher p_v wins. Hence, when $p_v > p_{v'}$, the strains with p_v (blue dots) are shown at a higher steady state. When $p_{v'} > p_v$, the red dots are shown at a higher steady state. (b) In the compartmental model, the strain with the higher carrying capacity wins. The ratio of p_v to $p_{v'}$ is no longer the factor that determines which strain wins.

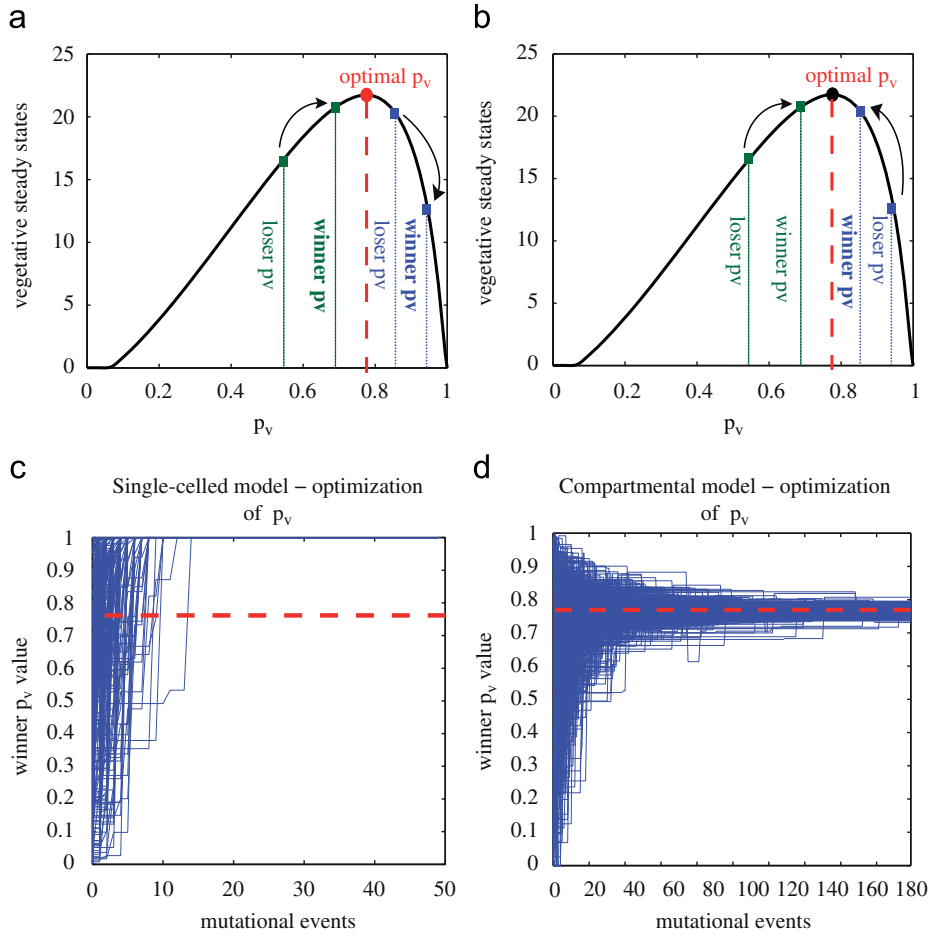


Fig. 5. Optimization of vegetative to heterocyst cell ratios. Plots (a) and (b) illustrate the winning factor in competitions between strains with different p_v. The parabola-shaped curve represents the carrying capacity as function of p_v. For each case, the carrying capacity of two pairs of strains simulated in isolation are shown as examples (green and blue squares). Next to each strain, we indicate the outcome of the competition within the pair (winner or loser). (a) In the single-celled model, the strain with the higher p_v wins. (b) In the compartmental model, the one reaching the higher carrying capacity wins. Plots (c) and (d) show the outcome of repeated mutational events, at which a competition between resident and mutant strain take place. (c) In the single-celled model, the optimal p_v (red line) is surpassed and $p_v \to 1$. (d) In the compartmental model, p_v tends to the optimal value.

not play a role anymore. In the multicellular organization, the winning indicator is the vegetative steady state value that would be reached with the corresponding p_v or $p_{v'}$ value in isolation. This winning indicator is in essence with the carrying capacity of each strain when grown separate from the other. Given any two competing strains, we found that the winning strain almost always had the higher carrying capacity when grown in isolation. This in turn depends on the corresponding proportion of vegetative and heterocyst cells produced during reproduction. Given equal initial conditions for both compartments, the carrying capacity holds true as the winning factor in 97–98% of cases when the full parameter space $\mathbb{R}^{22}$ is sampled randomly (details in Supplementary Information). The 2–3% exceptions correspond to cases in which (i) a numerical error occurs; (ii) $p_v \simeq p_{v'}$, hence the time required for the populations to stabilize is longer than the simulation time; (iii) the p_v value of the loser is too close to 0, hence the corresponding population can not grow.

3.2.1. The optimal rate of differentiation (p_v)

The vegetative cell steady state is dependent on the p_v value. We found that the carrying capacity for the number of vegetative cells is a parabola-shaped function of p_v (Figs. 5a and b). Hence, a population too rich in vegetative cells would be disadvantaged in comparison to a population with p_v closer to the maximum of the curve. Identical optimality conditions hold in both models. Interestingly, the fact that the optimal p_v values are usually above 0.5 agrees with other theoretical work (Willensdorfer, 2009) indicating that the optimal fraction of germline cells in simple multicellular organisms will be higher than that of somatic cells. It also agrees with the high proportion of vegetatives seen in cyanobacteria. Figs. 5a and b illustrate further what we see in Figs. 4a and b. Consider the outcome of competitions between a pair of strains with different p_v values. In the single-celled model, the strain with the higher p_v wins, hence the optimal p_v can be surpassed (Fig. 5a). In the compartmental model, the population with the potential of a higher carrying capacity outcompetes the other, getting closer to the optimal p_v (Fig. 5b). These results led us to consider the case of repeated competitions, as analyzed in the following section.

3.2.2. Evolutionary optimization of p_v

Stochastic simulations of successive mutational events test the ability of the two strategies to evolve towards the optimal p_v value. Fig. 5c shows the outcome in the single-celled model. The population always evolves to a full cheater situation, where $p_v = 1$. Hence in this case optimization is not possible. After each mutational event the population with the higher p_v value will go to fixation, hence increasing p_v towards 1. Fig. 5d shows the results of the compartmental model. In this case, mutant competitions automatically lead to optimization towards the p_v value that corresponds to the maximum steady state value of vegetatives. Hence, compartmentalization allows populations to evolve towards optimal ratios of vegetative to heterocyst cells.

3.3. Duration of daylight and separation of tasks in time and space

We numerically solve systems (5) and (10) separately. Fig. 6b shows the steady states of the populations going from permanent darkness to permanent light. In the extreme case of permanent darkness, both species die because of the lack of photosynthesis. In regimes where the percentage of daylight is scarce, the species following the external day/night periodicity reaches a higher carrying capacity than the multicellular differentiated species. The opposite happens when the duration of daylight is much longer

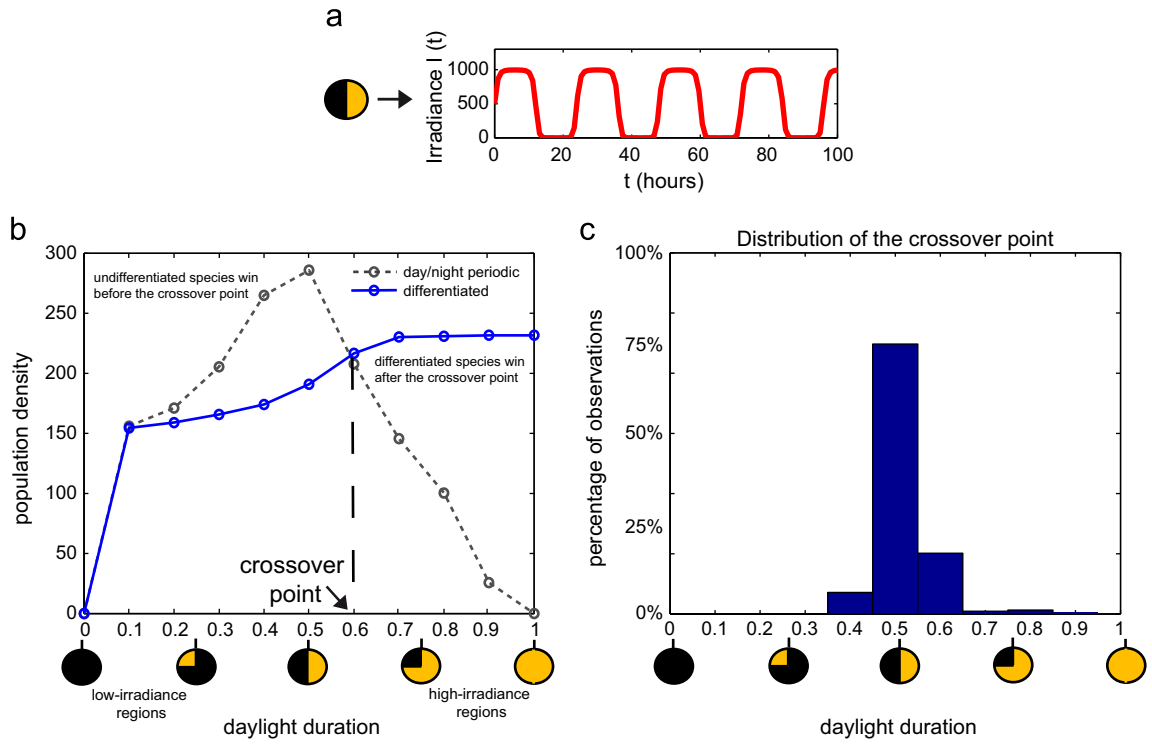


Fig. 6. (a) The irradiance function $I(t)$ as in Eq. (7), when the duration of day and night are equal. (b) Case study showing the comparison of the steady states of species following the day/night periodicity vs. differentiated species for different external irradiance cycles. Undifferentiated species that follow the day/night alternation reach a higher steady state in environments where the daylight is less or equal than the dark period. The point after which the differentiated species perform better is indicated by the crossover point. (c) Distribution of the crossover points based on 6000 trials. Parameters a, q, r, A were sampled randomly. In 75% of the cases, the crossover point locates around 0.5.

than night. In the latter case, as the daylight period is extended, the production of nitrogen decreases in the undifferentiated bacteria until it equals or is less than its consumption. In this situation the steady state population decays to 0. Hence long daylight conveys an advantage to division of tasks in space by means of differentiation. Fig. 6b shows a crossover point between the steady states, after which periodic species perform worse than the heterocystous. In order to check the occurrence and the location of a crossover point for a more general parameter space, we repeated 6000 times the scan of external daylight percentage for the two models. Each time we sampled randomly the values of parameters a, q, r, A. We recorded the position of the crossover point and we plotted the corresponding distribution through the histogram showed in Fig. 6c. We can conclude that in the 98.8% of the cases, there was always a crossover point after which the undifferentiated species reach a lower carrying capacity than the differentiated species. In the majority of the cases (75% of the cases), the crossover point is at 0.5, corresponding to a situation where the duration of night equals that of the day. These results indicate that environments where the dark period is significantly longer than the daylight period can be disadvantageous to terminally differentiated species.

3.4. Phylogenetic relationships among cyanobacteria

Phylogenetic relationships from 16S rRNA sequence of 37 strains of cyanobacteria are shown in Fig. 7. Cyanobacterial species were grouped into classes I–V as described by Rippka et al. (1979). Only the Bayesian tree is shown. Though the topology is based on

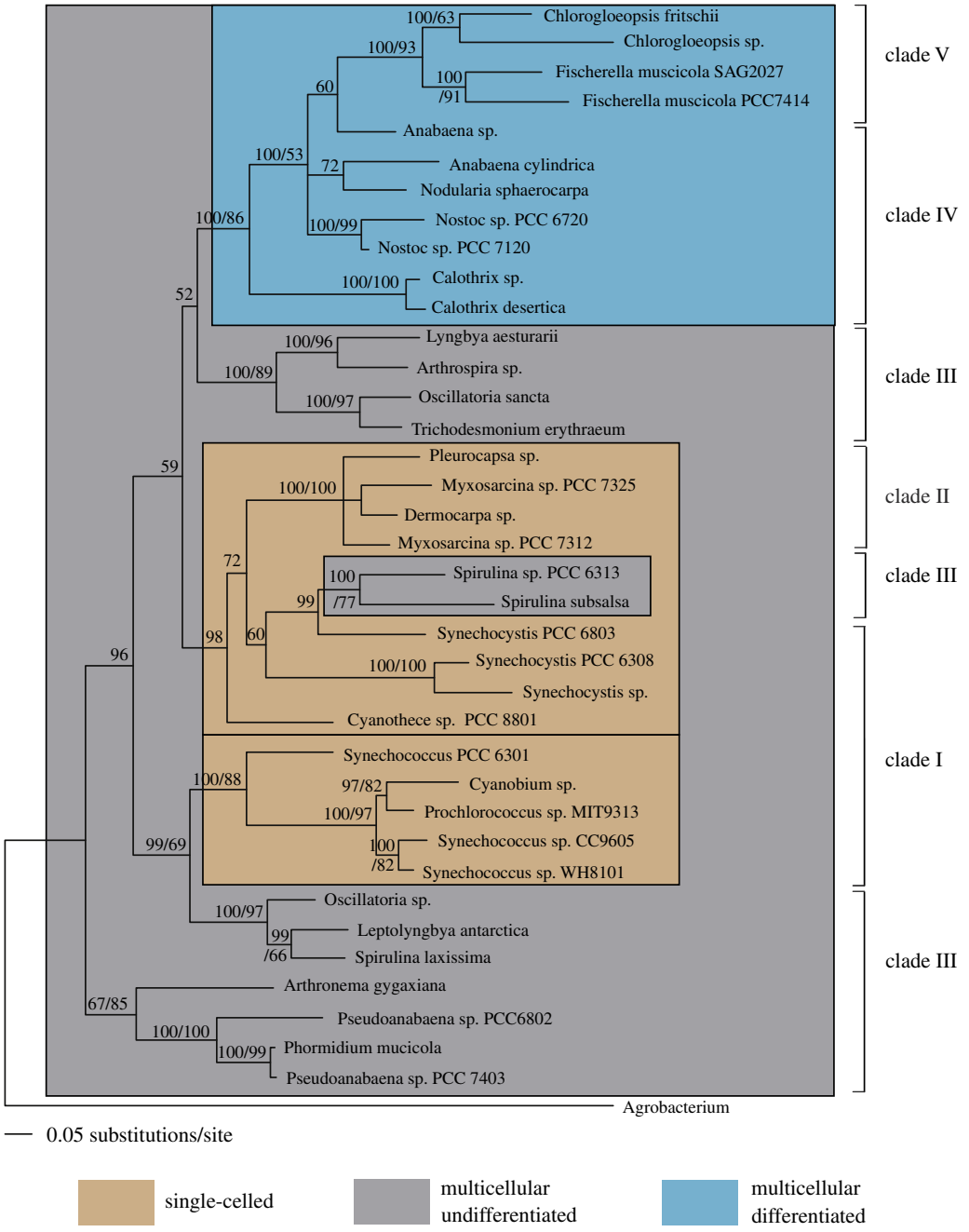


Fig. 7. Phylogenetic tree of cyanobacteria. Bayesian-tree of 16S rRNA sequences, based on GTR+I+G substitution model, with *Agrobacterium tumefaciens* as an outgroup. Shown at the nodes are only posterior probabilities (> 50%) or both posterior probabilities/bootstrap (> 50%). Posterior probabilities were calculated from 12,001 trees. Bootstrap values are obtained from 400 pseudo-replicates with maximum likelihood.

Bayesian analysis, character states are colored according to parsimony criteria (such that the least changes occur along the branches). Our analyses confirm the polyphyly of single celled clade I and the multicellular clade III as reported earlier (Giovannoni et al., 1988; Seo and Yokota, 2003). The multicellular, terminally differentiated clades IV and V, together form a monophyletic group supported by posterior probability (100%) and bootstrap (86%). The latter monophyly has been reported by other studies (Turner et al., 1999; Garcia-Pichel et al., 2001; Tomitani et al., 2006). Species from the polyphyletic clade III, belonging to the genera *Lyngbya*, *Arthrospira*, *Oscillatoria* and *Trichodesmium*, form the sister group of the monophyletic clade IV and V (blue box in Fig. 7). The phylogeny supports the conclusions of the simulations, according to which undifferentiated multicellularity evolved first, and hence made terminal differentiation possible.

4. Discussion

At first glance, multicellularity can appear as an obvious prerequisite for cellular differentiation. However, from a logical perspective, alternative developmental strategies are in principle possible. It has been recently emphasized (Leimar and Hammerstein, 2006) that it is important to strengthen the connection between theoretical models on the evolution of cooperation and explicit empirical cases. The framework we present here is formulated with this goal in mind, whereby we take a mechanistic representation of known biochemical interactions in an important group of organisms (the cyanobacteria), and show how the latter interactions fit into theoretical frameworks that attempt to explain multicellularity and the division of labor.

As discussed in Section 1.3, practically all solutions for avoiding the tragedy of the commons involve somehow separating the population into competing subsets. The results in Sections 3.1 and 3.2 are no exception to the latter rule. Compartmentalization allows for the protection of vital resources from potential disruptive mutations, whose effect can be limited to the compartment they arise in. Furthermore, multicellularity guarantees that cells in a compartment are clones. One could in fact make the argument that the evolution of multicellularity—and thereby compartmentalization—is a mechanistic means by which kin selection becomes "hard-wired" for a population of cooperating cyanobacterial cells. The cell interactions in the cyanobacterial system also have some structural similarities to a two component hypercycle with a single self-replicating catalyst, the main difference being the lack of hypergeometric growth in the bacterial case. It is noteworthy that conclusions previously derived from the study of hypercycles (Eigen and Schuster, 1978; Michod, 1983; Szathmáry and Demeter, 1987) can also apply to cell interactions and the evolution of multicellularity, indicating the potential generality of the former theory.

The cyanobacterial cell system has also some commonalities with cooperation games. The tragedy of the commons can for example be characterized by games like Prisoner's Dilemma, where the optimal strategy corresponds to cooperation of both players. However, the cell interactions that we consider do not directly map to a simple n-player game with a payoff matrix. Nonetheless, one may say that populations that converge on the optimal p_v, are in a state where all individuals are cooperating.

As seen in Section 3.2, the vegetative/heterocyst ratio has an effect on the carrying capacity of the population. Hence the tuning of the proportion of vegetatives upon division (p_v) can lead to the maximization of the carrying capacity. The autonomous optimization of the carrying capacity after repeated mutational events is found to be very different in the two models, as shown in Section 3.2.2. The single-celled strategy causes the fixation of the variant producing the most vegetative cells, thus converging to the value corresponding to the pure cheater case. Hence, higher fertility in the short term leads to decrease of carrying capacity and eventual extinction of the population in the long term. This explains why this evolutionary step is not observed in nature (see Fig. 8). On the other hand, the multicellular strategy allows for optimization towards the most profitable proportion of vegetative and heterocyst cells and for the selection and fixation of mutants that correspond to the maximal carrying capacity achievable by the population. Interestingly, this observation agrees with the almost constant vegetative/heterocysts ratio seen in many species of filamentous cyanobacteria (Adams, 2000).

The results from Sections 3.1 and 3.2 exclude the possibility of a transition from the undifferentiated unicellular stage to a differentiated single-celled one. Therefore, to achieve division of labor in cyanobacteria, two other paths are in principle possible: from undifferentiated unicellularity directly to differentiated multicellularity, or via the intermediate step of undifferentiated multicellularity (Fig. 8). The outcome of the phylogenetic analyses in Section 3.4 supports the second alternative, providing empirical evidence that the route to division of labor has included undifferentiated multicellularity. The combination of theoretical and phylogenetic results presented here lead to the conclusion that for the class of interactions occurring in cyanobacteria, multicellularity is a necessary condition for the evolution of terminal differentiation and the optimization of division of labor.

To further understand the ecological factors affecting the evolution of differentiation, we compared the advantages of a spatial separation of tasks over a temporal separation. The geographic distribution of cyanobacteria varies from mild to extreme environments (Paerl et al., 2000). It is known that environmental factors such as temperature can favor different forms of differentiation in cyanobacteria (Staal et al., 2003). However, there is at present no clear understanding of the distribution pattern of differentiated vs. undifferentiated cyanobacteria. According to results from our model of bacteria following the day/night periodicity in Section 3.3, division of labor by means of terminal differentiation is advantageous when the proportion of day is higher than that of night. In the latter case, the undifferentiated cyanobacteria fix nitrogen only in the short dark period. Meanwhile heterocystous cyanobacteria fix nitrogen also during the long day period. Conversely, in cases with scarce daylight, undifferentiated species have an advantage because during the short daylight period all available cells are devoted to light harvesting.

The undifferentiated species that we model simply follow the day/night alternation imposed by the external conditions, without the possibility of development of an internal cycle. The evolution of a self-regulated cycle that is independent of external light

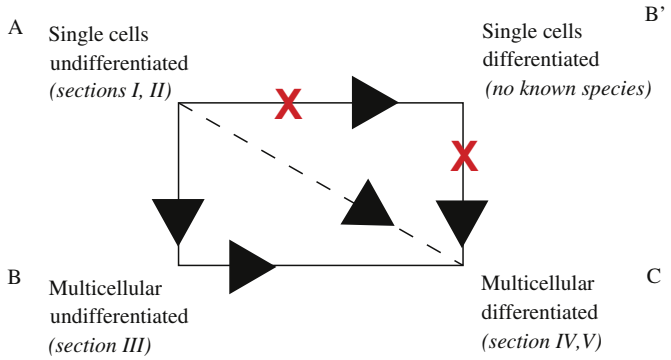


Fig. 8. The evolutionary paths leading to division of labor in cyanobacteria. A direct transition from A to C is in principle possible. Simulations exclude node B′. The phylogenetic tree supports the path B→C.

periodicity can potentially enhance the fitness of such bacteria, because the cycle can then be optimized according to the resource requirements. Hence if the circadian rhythm is optimized, there is the possibility that the undifferentiated circadian species can be also competitive in regions with long daylight periods. A true circadian rhythm has been observed in unicellular cyanobacteria such as *Synechococcus* (Mitsui et al., 1986). Further investigation of the benefits provided by an internal cycle could give an explanation for the maintenance of circadian rhythms in cyanobacteria. On the other hand, the development and regulation of such complex mechanisms is costly for the organism, and one may argue that a high cost of switching could support selection for differentiated species.

Our results on the response to daylight periodicities provide the general conclusion that in an environment with a short light period, selection acts against heterocystous cyanobacteria. In regimes of prevailing darkness, the absence of differentiation and the evolution of a circadian rhythm—or at least the simple adjustment according to the external periodicity—is advantageous. Hence adaptation to long daylight periods can be indicated as a possible reason for the evolution of terminal differentiation in cyanobacteria. The latter is a hypothesis that can be subject to empirical testing. Laboratory experiments can determine the outcome of competitions between undifferentiated and differentiated species under different day/night regimes. In addition, in order to determine seasonal differences between differentiated and undifferentiated species, ecological observations involving sample collection and quantitative measurements of species abundances could be systematically done in different seasons and at different latitudes.

Acknowledgments

We thank Andrew Barbour and two anonymous reviewers for suggestions that considerably improved this work. We thank Martin Ackermann, Sebastian Bonhoeffer, Leo Buss, Marta Manser and Katarzyna Palinska for many helpful comments. This project was supported by *Kantonal* and *Forschungskredit* funds of the University of Zurich.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi: 10.1016/j.jtbi.2009.09.009

References

Ackermann, M., Stecher, B., Freed, N.E., Songhet, P., Hardt, W.-D., Doebeli, M., 2008. Self-destructive cooperation mediated by phenotypic noise. Nature 454 (7207), 987–990.

Adams, D.G., 2000. Heterocyst formation in cyanobacteria. Current Opinion in Microbiology 3 (6), 618–624.

Agawin, N.S.R., Rabouille, S., Veldhuis, M.J.W., Servatius, L., Hol, S., van Overzee, H.M.J., Huisman, J., 2007. Competition and facilitation between unicellular nitrogen-fixing cyanobacteria and non-nitrogen-fixing phytoplankton species. Limnology and Oceanography 52 (5), 2233–2248.

Attolini, C.S.O., Stadler, P.F., 2006. Evolving towards the hypercycle: a spatial model of molecular evolution. Physica D—Nonlinear Phenomena 217 (2), 134–141.

Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. Science 211 (4489), 1390–1396.

Bach, L.A., Helvik, T., Christiansen, F.B., 2006. The evolution of n-player cooperation—threshold games and ESS bifurcations. Journal of Theoretical Biology 238, 426.

Bergman, B., Gallon, J.R., Rai, A.N., Stal, L.J., 1997. N-2 fixation by non-heterocystous cyanobacteria. FEMS Microbiology Reviews 19 (3), 139–185.

Boerlijst, M.C., Hogeweg, P., 1991. Spiral wave structure in pre-biotic evolution—hypercycles stable against parasites. Physica D 48 (1), 17–28.

Bonner, J.T., 2000. First Signals: The Evolution of Multicellular Development. Princeton University Press, Princeton.

Boomsma, J.J., Franks, N.R., 2006. Social insects: from selfish genes to self organisation and beyond. Trends in Ecology and Evolution 21 (6), 303–308.

Brockhurst, M.A., Buckling, A., Racey, D., Gardner, A., 2008. Resource supply and the evolution of public-goods cooperation in bacteria. BMC Biology 6.

Buss, L.W., 1988. The Evolution of Individuality. Princeton University Press, Princeton, NJ.

Buss, L.W., 1982. Somatic-cell parasitism and the evolution of somatic tissue compatibility. Proceedings of the National Academy of Sciences of the United States of America—Biological Sciences 79 (17), 5337–5341.

Buss, L.W., 1983. Evolution, development, and the units of selection. Proceedings of the National Academy of Sciences of the United States of America—Biological Sciences 80 (5).

Doebeli, M., Knowlton, N., 1998. The evolution of interspecific mutualisms. Proceedings of the National Academy of Sciences 95 (15), 8676–8680.

Eigen, M., 1971. Selforganization of matter and evolution of biological macromolecules. Naturwissenschaften 58 (10), 465.

Eigen, M., Gardiner, W.C., Schuster, P., 1980. Hypercycles and compartments—compartments assists—but do not replace—hypercyclic organization of early genetic information. Journal of Theoretical Biology 85 (3), 407–411.

Eigen, M., Schuster, P., 1977. Hypercycle—principle of natural self-organization. A. emergence of hypercycle. Naturwissenschaften 64 (11), 541–565.

Eigen, M., Schuster, P., 1978. Hypercycle—principle of natural self-organization. B. abstract hypercycle. Naturwissenschaften 65 (1), 7–41.

Enquist, M., Leimar, O., 1993. The evolution of cooperation in mobile organisms. Animal Behaviour 45 (4), 747–757.

Ferriere, R., Bronstein, J.L., Rinaldi, S., Law, R., Gauduchon, M., 2002. Cheating and the evolutionary stability of mutualisms. Proceedings of the Royal Society of London B 296, 773–780.

Ferriere, R., Michod, R.E., 1996. The evolution of cooperation in spatially heterogeneous populations. American Naturalist 147 (5), 692–717.

Flores, E., Herrero, A., Wolk, C.P., Maldener, I., 2006. Is the periplasm continuous in filamentous multicellular cyanobacteria?. Trends in Microbiology 14 (10), 439–443.

Fontanari, J.F., Santos, M., Szathmary, E., 2006. Coexistence and error propagation in pre-biotic vesicle models: a group selection approach. Journal of Theoretical Biology 239 (2), 247–256.

Frank, S.A., 1994. Kin selection and virulence in the evolution of protocells and parasites. Proceedings of the Royal Society of London Series B—Biological Sciences 258 (1352), 153–161.

Garcia-Pichel, F., Lopez-Cortes, A., Nuebel, U., 2001. Phylogenetic and morphological diversity of cyanobacteria in soil desert crusts from the Colorado Plateau. Applied and Environmental Microbiology 67 (4), 1902–1910.

Gilbert, O.M., Foster, K.R., Mehdiabadi, N.J., Strassmann, J.E., Queller, D.C., 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. Proceedings of the National Academy of Science of the United States of America 104 (21), 8913–8917.

Giovannoni, S.J., Turner, S., Olsen, G.J., Barns, S., Lane, D.J., Pace, N.R., 1988. Evolutionary relationships among cyanobacteria and green chloroplasts. Journal of Bacteriology 170 (8), 3584–3592.

Griffin, A.S., West, S.A., Buckling, A., 2004. Cooperation and competition in pathogenic bacteria. Nature 430 (7003), 1024–1027.

Hamilton, W.D., 1964a. The genetical evolution of social behaviour. I. Journal of Theoretical Biology 7 (1), 1.

Hamilton, W.D., 1964b. The genetical evolution of social behaviour. II. Journal of Theoretical Biology 7 (1), 17.

Hammerstein, P., 2003. Why is reciprocity so rare in social animals? A protestant appeal. In: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation, Dahlem Workshop Reports: Environmental Sciences Research Report, 90th Dahlem Workshop on Genetic and Cultural Evolution of Cooperation, Berlin, Germany, June 23–28, 2002, pp. 83–93.

Han, B.P., 2001. Photosynthesis-irradiance response at physiological level: a mechanistic model. Journal of Theoretical Biology 213 (2), 121.

Hardin, G., 1968. The tragedy of the commons. Science 162, 143.

Hochberg, M., Rankin, D., Taborsky, M., 2008. The coevolution of cooperation and dispersal in social groups and its implications for the emergence of multicellularity. BMC Evolutionary Biology 8 (1), 238.

Hofbauer, J., Schuster, P., Sigmund, K., 1979. Evolutionary stable strategies and game dynamics. Journal of Theoretical Biology 81 (3), 609–612.

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge.

Hogeweg, P., Takeuchi, N., 2003. Multilevel selection in models of prebiotic evolution: compartments and spatial self-organization. Origins of Life and Evolution of Biospheres 33 (4), 375.

Honda, D., Yokota, A., Sugiyama, J., 1999. Detection of seven major evolutionary lineages in cyanobacteria based on the 16s rRNA gene sequence analysis with new sequences of five marine Synechococcus strains. Journal of Molecular Evolution 48 (6), 723–739.

Kageyama, H., Nishiwaki, T., Nakajima, M., Iwasaki, H., Oyama, T., Kondo, T., 2006. Cyanobacterial circadian pacemaker: Kai protein complex dynamics in the KaiC phosphorylation cycle in vitro. Molecular Cell 23 (2), 161–171.

Kaiser, D., 2001. Building a multicellular organism. Annual Review of Genetics 35 (1), 103–123.

Killingback, T., Bieri, J., Flatt, T., 2006. Evolution in group-structured populations can resolve the tragedy of the commons. Proceedings of the Royal Society B—Biological Sciences 273 (1593), 1477–1481.
Kondo, T., Strayer, C.A., Kulkarni, R.D., Taylor, W., Ishiura, M., Golden, S.S., Johnson, C.H., 1993. Circadian rhythms in prokaryotes—luciferase as a reporter of circadian gene expression in cyanobacteria. Proceedings of the National Academy of Sciences of the United States of America 90 (12), 5672–5676.
Kurosawa, G., Aihara, K., Iwasa, Y., 2006. A model for the circadian rhythm of cyanobacteria that maintains oscillation without gene expression. Biophysical Journal 91 (6), 2015–2023.
Lawlor, D.W., 1990. Photosynthese. Georg Thieme Verlag, Stuttgart.
Lehmann, L., Keller, L., 2006. The evolution of cooperation and altruism—a general framework and a classification of models. Journal of Evolutionary Biology 19 (5), 1365–1376.
Leimar, O., Hammerstein, P., 2001. Evolution of cooperation through indirect reciprocity. Proceedings of the Royal Society of London Series B—Biological Sciences 268 (1468), 745–753.
Leimar, O., Hammerstein, P., 2006. Facing the facts. Journal of Evolutionary Biology 19 (5), 1403–1405.
Mariscal, V., Herrero, A., Flores, E., 2007. Continuous periplasm in a filamentous heterocyst-forming cyanobacterium. Molecular Microbiology 65 (4), 1139–1145.
Marler, M.R., Gehrman, P., Martin, J.L., Ancoli-Israel, S., 2006. The sigmoidally transformed cosine curve: a mathematical model for circadian rhythms with symmetric non-sinusoidal shapes. Statistics in Medicine 25 (22), 3893–3904.
Michod, R.E., 2007. Evolution of individuality during the transition from unicellular to multicellular life. Proceedings of the National Academy of Sciences 104 (Suppl. 1), 8613–8618.
Michod, R.E., 1983. Population biology of the 1st replicators—on the origin of the genotype, phenotype and organism. American Zoologist 23 (1), 5–14.
Michod, R.E., Roze, D., 2001. Cooperation and conflict in the evolution of multicellularity. Heredity 86 (Part 1), 1–7.
Michod, R.E., Viossat, Y., Solari, C.A., Hurand, M., Nedelcu, A.M., 2006. Life-history evolution and the origin of multicellularity. Journal of Theoretical Biology 239.
Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S., Arai, T., 1986. Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. Nature 323 (6090), 720–722.
Mullineaux, C.W., Mariscal, V., Nenninger, A., Khanum, H., Herrero, A., Flores, E., Adams, D.G., 2008. Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria. EMBO Journal 27 (9), 1299–1308.
Nakamaru, M., Matsuda, H., Iwasa, Y., 1997. The evolution of cooperation in a lattice-structured population. Journal of Theoretical Biology 184 (1), 65–81.
Nowak, M.A., Bonhoeffer, S., May, R.M., 1994. Spatial games and the maintenance of cooperation. Proceedings of the National Academy of Sciences of the United States of America 91 (11), 4877–4881.
Nowak, M.A., May, R.M., 1992. Evolutionary games and spatial chaos. Nature 359 (6398), 826–829.
Ohtsuki, H., Hauert, C., Lieberman, E., Nowak, M.A., 2006. A simple rule for the evolution of cooperation on graphs and social networks. Nature 441 (7092), 502–505.
Paerl, H.W., Pinckney, J.L., Steppe, T.F., 2000. Cyanobacterial-bacterial mat consortia: examining the functional unit of microbial survival and growth in extreme environments. Environmental Microbiology 2 (1), 11–26.
Partensky, F., Hess, W.R., Vaulot, D., 1999. Prochlorococcus, a marine photosynthetic prokaryote of global significance. Microbiology and Molecular Biology Reviews 63 (1), 106–127.
Pfeiffer, T., Bonhoeffer, S., 2003. An evolutionary scenario for the transition to undifferentiated multicellularity. Proceedings of the National Academy of Sciences 100 (3), 1095–1098.
Pfeiffer, T., Schuster, S., Bonhoeffer, S., 2001. Cooperation and competition in the evolution of ATP-producing pathways. Science 292 (5516), 504–507.
Rabouille, S., Staal, M., Stal, L., Soetaert, K., 2006. Modeling the dynamic regulation of nitrogen fixation in the cyanobacterium Trichodesmium sp. Applied and Environmental Microbiology 72 (5), 3217–3227.
Rainey, P.B., Rainey, K., 2003. Evolution of cooperation and conflict in experimental bacterial populations. Nature 425 (6953), 72–74.
Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., Stanier, R.Y., 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Journal of General Microbiology 111, 1–61.
Roberts, G., Sherrat, T.N., 1998. Development of cooperative relationships through increasing investment. Nature 394, 175–179.
Roussel, M.R., Gonze, D., Goldbeter, A., 2000. Modeling the differential fitness of cyanobacterial strains whose circadian oscillators have different free-running periods: comparing the mutual inhibition and substrate depletion hypotheses. Journal of Theoretical Biology 205 (2), 321–340.
Rubio, F.C., Camacho, F.G., Sevilla, J.M.F., Chisti, Y., Grima, E.M., 2003. A mechanistic model of photosynthesis in microalgae. Biotechnology and Bioengineering 81 (4), 459–473.
Sachs, J.L., Mueller, U.G., Wilcox, T.P., Bull, J.J., 2004. The evolution of cooperation. The Quarterly Review of Biology 79 (2), 135–160.
Saier, M., Jacobson, G.R., 1984. The Molecular basis of Sex and Differentiation. Springer, New York.
Sardanyés, J., Solé, R.V., 2006. Bifurcations and phase transitions in spatially extended two-member hypercycles. Journal of Theoretical Biology 243 (4), 468–482.
Schleip, W., 1934. August Weismanns Bedeutung für die Entwicklung der Zoologie und allgemeinen Biologie zu seinem hundertsten Geburtstag am 17. Januar 1934. Naturwissenschaften 22 (3).
Schopf, J.W., 1994. Disparate rates, differing fates: tempo and mode of evolution changed from the Precambrian to the Phanerozoic. Proceedings of the National Academy of Sciences 91, 6735–6742.
Seo, P.S., Yokota, A., 2003. The phylogenetic relationships of cyanobacteria inferred from 16srRNA, gyrB, rpoC1 and rpoD1 gene sequences. The Journal of General and Applied Microbiology 49.
Shapiro, J.A., 1998. Thinking about bacterial populations as multicellular organisms. Annual Review of Microbiology 52 (1), 81–104.
Smith, J.M., 1964. Group selection and kin selection. Nature 201, 1145–1147.
Smith, J.M., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.
Smith, J.M., 1979. Hypercycles and the origin of life. Nature 280 (5722), 445–446.
Smith, J.M., Price, G.R., 1973. Logic of animal conflict. Nature 246 (5427), 15–18.
Smith, J.M., Szathmáry, E., 1995. The Major Transitions in Evolution. Freeman, Oxford.
Staal, M., Meysman, F.J.R., Stal, L.J., 2003. Temperature excludes N-2-fixing heterocystous cyanobacteria in the tropical oceans. Nature 425 (6957) 504–507.
Stal, L.J., Krumbein, W.E., 1987. Temporal separation of nitrogen-fixation and photosynthesis in the filamentous, nonheterocystous cyanobacterium Oscillatoria sp. Archives of Microbiology 149 (1), 76–80.
Strassmann, J.E., Zhu, Y., Queller, D.C., 2000. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. Nature 408 (6815), 965–967.
Suzuki, S., Akiyama, E., 2008. Chaos, oscillation and the evolution of indirect reciprocity in n-person games. Journal of Theoretical Biology 252, 686.
Szathmáry, E., Demeter, L., 1987. Group selection of early replicators and the origin of life. Journal of Theoretical Biology 128 (4), 463–486.
Tilman, D., 1977. Resource competition between planktonic algae—experimental and theoretical approach. Ecology 58 (2), 338–348.
Tomitani, A., Knoll, A.H., Cavanaugh, C.M., Ohno, T., 2006. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. Proceedings of the National Academy of Sciences of the United States of America 103 (14), 5442–5447.
Travisano, M., Velicer, G.J., 2004. Strategies of microbial cheater control. Trends in Microbiology 12 (2), 72–78.
Trivers, R.L., 1971. Evolution of reciprocal altruism. Quarterly Review of Biology 46 (1), 35.
Turner, S., Pryer, K.M., Miao, V.P.W., Palmer, J.D., 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small submit rRNA sequence analysis. Journal of Eukaryotic Microbiology 46 (4), 327–338.
Velicer, G.J., Kroos, L., Lenski, R.E., 2000. Developmental cheating in the social bacterium *Myxococcus xanthus*. Nature 404 (6778), 598.
Vrede, K., Heldal, M., Norland, S., Bratbak, G., 2002. Elemental composition (C, N, P) and cell volume of exponentially growing and nutrient-limited bacterioplankton. Applied and Environmental Microbiology 68 (6), 2965–2971.
Weismann, A., 1889. Essays Upon Heredity. Oxford At the Clarendon Press (Chapters II, IV) ⟨http://www.esp.org:80/books/weismann/essays/facsimile/⟩.
West, S.A., Griffin, A.S., Gardner, A., Diggle, S.P., 2006. Social evolution theory for microorganisms. Nature Reviews Microbiology 4 (8), 597–607.
Willensdorfer, M., 2009. On the evolution of differentiated multicellularity. Evolution 63 (2), 306–323.
Wolk, C.P., 1996. Heterocyst formation. Annual Review of Genetics 30, 59–78.

Acknowledgements

Many people have contributed in various ways to the completion of my dissertation and it is a pleasure to thank those who made this thesis possible.

I am deeply grateful to my supervisor **Homayoun C. Bagheri**. It was an intellectual pleasure working with him. During the last four years he has supported me in whatever way he could. He has constantly encouraged me to work independently and integrate my own ideas into the project. Thanks to his assistance and good advice this dissertation could be successfully completed. Stimulating conversations and excellent mentoring has prepared me for a confident start into an academic career.

Furthermore, I would like to show my gratitude to the members of my research committee: Interesting discussions and comments on my work have facilitated the finish of my studies. Thank you, **Kentaro K. Shimizu** and **Lukas Keller**.

My dissertation resulted in five publications. I would like to thank the authors of two successful publications in which I enjoyed to contribute as a coauthor: Valentina Rossetti, Marco Bernasconi, Christoph Sandrock and Christoph Vorburger. Thanks to Marco Bernasconi especially for the patience of introducing me to the field of phylogenetics.
Furthermore, I am grateful for the help on three of my first-author publications. I would like to show my gratitude to Maria Anisimova and Daniel A. Dalquen for straightforward and elaborate team work. Working with Alexandre Antonelli was a pleasure during which I was able to accomplish various new skills. My regards go further to my enthusiastic colleague Jurriaan de Vos for the highly efficient collaboration on my fourth chapter.

A special thanks goes also to the members of the group of Evolutionary Genetics and Theoretical Biology for their constructive suggestions and support in the lab: Valentina Rossetti, Manuela Filippini Cattani, Miroslav Svercel, Akos Dobay, Mahtab Mohebbi and Firouzeh Farahmand.

For helpful comments on my manuscripts I would like to thank Elena Conti and Bryan Moore. I am also grateful for excellent statistical advice from Erik Postma. For strong encouragement as well as for superb advice on application proposals I am indebted to Barbara Tschirren.

I would like to thank every single the members of the Institute of Evolutionary Biology and the Faculty of Science at the University of Zurich. Working with so many bright scientists in such a friendly and open minded atmosphere was a delight. Some colleagues I would like to thank separately for helpful advice in the past four years: thank you, Heinz-Ulrich

Reyer, Bernhard Schmid, Jakob Pernthaler, Tony Wilson, Helmut Brandl, Josh Van Buskirk and Pamela Alean-Kirkpatrick.
For enormous social and administrative support I am grateful to Anni Maeder, Isabel Schöchli and Susanne Bischof.
For computer-related help I very much thank Tina Siegentaler and Michel Nakano.
Many colleagues and at the same time good friends would give me valuable scientific feedback or help me in another way. I would like to thank Kirsty Agnoli, Nalini Puniamoorthy, Anthony Palladino, Alexandra Hoffmann and Yann Hautier.

I owe my deepest gratitude to my mother, Christiane Schirrmeister for her unlimited support in so many ways it would fill pages if I would go into details. Furthermore, I would like to thank my father, Volker Schirrmeister, who served as a scientific role model and my brother Ralph Schirrmeister whose praise -possibly not always justified- motivated me ever again. I am also most grateful to my grandparents Röschen and Bruno Wittusch which contributed in many ways to my development.

Various friends earn my gratitude for immense social support in sometimes tough times. Big thanks to: Marina, Petra, Kerstin, Suresh, Paulin, Matthias, Ralli, Jens, Zuni, Romain, Matt, Nicole, Rapha, Erika, Adele, James and the palaeontology guys and all my friends from Nuremberg.

Most importantly, for his love, his encouragement, and support, I would like to thank **Christopher Reinmüller** with all my heart.

Bettina E. Schirrmeister, PhD student

CONTACT INFORMATION	Evolutionary Genetics and Theoretical Biology Institute of Evolutionary Biology and Environmental Studies University of Zurich Winterthurerstr. 190, 8057 Zurich, Switzerland	bettina.schirrmeister@ieu.uzh.ch **(office)** +41-44-6354738 **(mobile)** +41-76-7372930 www.ieu.uzh.ch/staff/phd/schirrmeister.html
CITIZENSHIP	German	DOB: 03 February 1981
OBJECTIVE	Postdoctoral research project	
RESEARCH INTERESTS	Phylogenetics, Evolutionary Genetics, Population Genetics, Bioinformatics, Palaeontology	

EDUCATION

7/2007 - today	**University of Zurich**, Zurich, Switzerland Doctoral candidate in Evolutionary Genetics, Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland Thesis Topic: *The Evolution of Morphological Complexity in Cyanobacteria* Advisor: Professor Homayoun C. Bagheri
10/2006 - 3/2007	**Max-Planck-Institute for Evolutionary Anthropology**, Leipzig, Germany Graduate research assistant in behavioral science: studying chimpanzees in Budongo Forest, Uganda, Africa
10/2005 - 9/2006	**Ludwig-Maximilians University of Munich**, Munich, Germany M.Sc. ("Diplom") in Population Genetics, Institute of Zoology, University of Munich, Germany Thesis Topic: *Searching for evidence of natural selection on the X-chromosome of European populations of Drosophila melanogaster* Advisor: Professor Wolfgang Stephan
10/2003 - 9/2005	**Ludwig-Maximilians University of Munich**, Munich, Germany Evolutionary Biology, Population Genetics, Anthropology, Zoology and Palaeontology
10/2000 - 9/2003	**Justus-Liebig University of Giessen**, Giessen, Germany Undergraduate Studies ("Vordiplom") in Biology

AWARDS AND HONORS

2011	**Best Poster** at Biology2011-conference. February 3rd-4th, Zurich, Switzerland
2010	**Best Talk** at PhD Meeting of Evolutionary Biology of the German-Zoological-Society. June 7th-9th, Freiburg im Breisgau, Germany

WORK EXPERIENCE

	University of Zurich, Zurich, Switzerland
12/2009 - 7/2011	Representative of PhD-students at the Institute for Evolutionary Biology and Environmental Studies
10/2009 - 7/2011	Member of the organizing committee for Graduation Ceremonies, Faculty of Science
11/2008 - 7/2011	Representative of PhD-students and research assistants at the Faculty of Science
7/2008 - 9/2010	Representative of PhD-students in the commission for biomedical dissertations across faculties
10/2009	Darwin-year 09, exhibition from the Zoological Institute at Zurich main station, Switzerland. Organizer and tour guide on *"The origin of life"*
4/2008	Exhibition for the 175 year Anniversary of the University of Zurich, Switzerland. Organization and talk on *"Natural Selection"*
	Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany
10/2006 - 3/2007	Graduate research assistant in behavioral science: studying chimpanzees in Budongo Forest, Uganda, Africa
	Ludwig-Maximilians University of Munich, Munich, Germany
4/2004 - 8/2005	Undergraduate research assistant, Evolutionary Biology, Institute of Zoology
	Justus Liebig University of Giessen, Giessen, Germany
7/2000-8/2000	Undergraduate laboratory assistant, Institute of General Botany

SEMINARS AND SUMMER SCHOOLS

8/2010 - 1 week	**University of Zurich**, Switzerland Course "Microbial Ecology: Biogeochemistry of Alpine Wetlands", Cadagno, Switzerland
5/2010 - 5 days	**Binais strategic services**, Switzerland Course "Management and Leadership - Summer school", Weggis, Switzerland
3/2008 - 2 days	**University of Zurich**, Switzerland Course "Debut-Introduction to teaching", Zurich, Switzerland
1/2008 - 10 days	**Ecole Normale Supérieure de Lyon**, France Course "Comparative Genomics", Lyon,France

SKILLS

Computing skills

OPERATING SYSTEMS	Linux, Windows, Mac OS X
GRAPHICS	Adobe illustrator, GIMP
DOCUMENT PREPARATION	LaTeX, various word processors and editors
SEQUENCE ANALYSIS	Bioedit, Clustal, PhyDe, Lasergene sequence analysis (SeqMan, EditSeq, MegAlign), MAFFT
PHYLOGENETICS	RAxML, MrBayes, GARLI, Mesquite, BayesTraits, Beast, Mega, TreeView, FigTree
STATISTICS	R

Lab skills

MOLECULAR	DNA-isolation, PCR, DNA-sequencing, Gel Electrophoresis
MICROBIOLOGICAL	Fieldwork experience, microscopy, cultivation of bacteria

LANGUAGES

NATIVE SPEAKER	German
FLUENT	English
BASIC	Spanish

MEMBERSHIPS

Swiss Systematics Society, the Swiss Red Cross, Touring Club Switzerland

INTERESTS

Travelling, archaeology, reading, hiking, high diving, soccer spectator

PRESENTATIONS

2011 **Talk** *The evolution of complexity in cyanobacteria*
Annual Research Symposium of the PhD-programm in Evolutionary Biology. July 4th-6th, Ascona, Switzerland

2011 **Poster** *The origin of multicellulrity in cyanobacteria*
Biology2011-conference. February 3rd-4th, Zurich, Switzerland

2010 **Talk** *The predominance of multicellularity in the early evolution of cyanobacteria*
Annual Research Symposium of the PhD-programm in Evolutionary Biology. June 7th-9th, Emmental, Switzerland

2010 **Talk** *The predominance of multicellularity in the early evolution of cyanobacteria*
PhD Meeting of Evolutionary Biology of the German-Zoological-Society. June 7th-9th, Freiburg im Breisgau, Germany

2009 **Talk** *The predominance of multicellularity in the early evolution of cyanobacteria*
15th European Meeting of PhD-Students in Evolutionary Biology "EMPSEB 09". August 14th-19th, Schoorl, The Netherlands

2009 **Talk** *The origin of cyanobacteria: Was the most recent common ancestor multicellular?*
Biology2009-conference. February 12th-13th, Bern, Switzerland

2008 **Talk** *The origin of cyanobacteria: Was the most recent common ancestor multicellular?*
Swiss Systematic Society Day. November 7th, Bern, Switzerland

MEDIA COVERAGE

2011 Initiated by **Schirrmeister et al. BMC Evol. Biol. 11:45**:

Science: Editor's Choice. *Shape-Shifting Cyanobacteria*, 332(6062):150

20 Minuten: *Evolution rückwärts*, March 24th

UZH News: *Vom Mehrzeller zum Einzeller und wieder zurück*, March 14th

The Scientist- Faculty of 1000: *Multicellular evolution not linear*, February 22nd

RESEARCH PUBLICATIONS

SUBMITTED

Schirrmeister BE*, Daquien D, Anisimova M, Bagheri HC. Gene copy number variation and its significance in cyanobacterial phylogeny (*submitted*)

Schirrmeister BE*, de Vos JM, Bagheri HC. Evolution to multicellularity coincided with diversification of cyanobacteria and the Great Oxidation Event (*submitted*)

PEER REVIEWED

2011 Sandrock C, Schirrmeister BE, Vorburger C. Evolution of reproductive mode variation and host associations in a sexual-asexual complex of aphid parasitoids. *BMC Evol. Biol.* **11:348**

2011 Schirrmeister BE*, Antonelli A, Bagheri HC (2011) The origin of multicellularity in cyanobacteria. *BMC Evol. Biol.* **11:45**. (***highly accessed***) *-ranked* #5 *of most viewed BMC Evol.Biol. articles in the past year*

2010 Rossetti V, Schirrmeister BE, Bernasconi MV, Bagheri HC (2010) The evolutionary path to terminal differentiation and division of labor in cyanobacteria. *Journal of Theoretical Biology* **262:23-34**.

INVITED ARTICLES

2011 Schirrmeister BE*, Anisimova M, Antonelli A, Bagheri HC (2011) Evolution of cyanobacterial morphotypes: Taxa required for improved phylogenomic approaches. *Communicative and Integrative Biology* **4(4):424-427**

*Corresponding author

REFERENCES

Homayoun C Bagheri
Professor
Evolutionary Genetics and Theoretical Biology
Institute of evolutionary Biology and Environmental Studies
University of Zurich, Switzerland
bagheri@ieu.uzh.ch
+41-44-6356623

Wolfgang Stephan
Professor
Population Genetics-Evolutionary Biology
Department Biology II
University of Munich, Germany
stephan@bio.lmu.de
+49-89-218074102

Alexandre Antonelli
PhD
Gothenburg Botanical Garden, Sweden
alexandre.antonelli@vgregion.se
+46-31-7411115

Erklärung

Hiermit bestätige ich, die vorliegende Dissertation selbständig und nur unter Verwendung der angegebenen Hilfsmittel und Quellen durchgeführt zu haben.

Zürich, Februar 2012 .....

Bettina E. Schirrmeister